# The Relationship Between Analogy And Categorisation In Cognition

**Michael Ramscar**

**Ph.D. University of Edinburgh 1999**

# Abstract

This central topic of this thesis is the relationship between categorisation and analogy in cognition. Questions of what a straightforward representation of a concept or category is, and following from that how extra-categorical associations such as analogy and metaphor are possible are central to our understanding of human reasoning and comprehension. However, despite the intimate linkage between the two, the trend in cognitive science has been to treat analogy and categorisation as separable, distinctive phenomena that can be studied in isolation from one another. This strategy has proved remarkably effective when it comes to the cognitive modelling of extra-categorical associations. A number of compelling and detailed models of analogy process exist, and there is widespread agreement amongst researchers studying analogy as to what the key cognitive processes that determine analogies are.

However, these models of analogy tend to assume some kind of fully specified category processing module which governs and determines ordinary, straightforward conceptual mappings. Indeed, this assumption is required in order to talk about analogy and metaphor in the first place: few theorists actually define analogy and metaphor per se, but all agree that analogical and metaphoric judgements can be defined in contrast to ordinary categorisation judgements.

This thesis reviews these models of analogy, and evidence for them, before conducting a detailed exploration of categorisation in relation to analogy. A theoretical and empirical review is presented in order to show that the straightforward notion of categorisation that underpins the distinctive phenomena approach to the study of analogy and categorisation is more apparent than real. Whilst intuitively, analogy and categorisation might feel like different things which can be contrasted with one another, from a cognitive processing point of view, this thesis argues that such a distinction may not survive a detailed scientific examination.

A series of empirical studies are presented in order to further explore the 'no distinction' hypothesis. Following from these, further studies examine the question of whether models of analogical processing have progressed as far as they can in artificial

isolation from categorisation, a process in which the processes that are normally deemed 'analogical' appear to play a vital role.

The conclusion drawn in this thesis is that the analogy / categorisation division, as currently formulated, cannot survive detailed scientific examination. It is argued that despite the benefits that the previous study of these phenomena in isolation have brought in the past, future progress, especially in the development of cognitive models of analogy, is dependent on a more unified approach.

# Declaration

I declare that this thesis is entirely my own composition, and that it describes my own research.

Michael Ramscar

Some of the work described in this thesis was published prior to submission. Experiment 1, described in Chapter 5, previously appeared as:

> Ramscar, M.J.A. and Pain, H.G., (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum Associates, New Jersey, pp 346-351.

Experiment 2, described in Chapter 5 appeared as:

> Ramscar, M.J.A., Pain, H.G., Darrington, S. and Lee, J. R. (1998) Examples and generalisations: Using surface versus structural recall biases to probe conceptual storage. in *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, Laurence Ealbaum Associates, pp 859-865.

Experiment 3, described in Chapter 5 was published as:

> Darrington, S., Lingstadt, T and Ramscar, M.J.A. (1998) Analogy as a sub-process of categorisation. in *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, Laurence Ealbaum Associates, pp 279-284.

Finally, analysis of analogical semantics in Chapter 6 appeared in condensed form as:

> Ramscar, M.J.A. Pain, H.G., & Cooper, R (1997) Is there a place for semantic similarity in the analogical mapping process? *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society.* Lawrence Earlbaum Associates., New Jersey, pp 632-636.

These papers are bound into this thesis in Appendix E. The following published papers also contain material that is included either in part or in greater detail in this thesis:

Ramscar. M.J.A., Lee, J.R. and Pain, H.G. (1996) A classification based methodology for the integration of agent views within design systems *Design Studies,* **17**, 4, pp 465 - 483,

includes an early version of the analogy review material presented in Chapter 2.

The following three papers:

Ramscar, M.J.A. (1997) Wittgenstein and the nature of psychological categories. in *Proceedings of SimCat 97*, Department of Artificial Intelligence Conference Proceedings, University of Edinburgh, Scotland, 205-211.

Ramscar, M.J.A. and Hahn, U. (1998a) Wittgenstein and the ontological status of analogy. In Holyoak, K, Gentner, D, and Kokinov, B *(eds.)* *Advances in Analogy Research*, New Bulgarian University Press, Sofia, Bulgaria, pp 390-402.

Ramscar, M.J.A. and Hahn, U. (1998b) What family resemblances are not. The continuing relevance of Wittgenstein in the study of concepts and categories. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society.* Madison-Wisconsin, USA, August, 1998, Laurence Ealbaum Associates, pp 865-870.

all include summaries of the exegesis of Wittgenstein's 'game' argument from Chapter 3, and Ramscar and Hahn (1998a) and Ramscar and Hahn (1998b) also contain material from the categorisation review in Chapter 4.

# Contents

## 4. CATEGORISATION II THEORIES AND MODELS OF CATEGORISATION ..84



## 5. PROBING THE "DISTINCTION" BETWEEN COGNITIVE THEORIES OF ANALOGY AND CATEGORISATION ........ 129

## 6. RE-EXAMINING THE ROLE OF SEMANTIC CONSTRAINTS IN ANALOGY162

# Acknowledgements

Firstly I wish to thank my family and friends, and especially Nik Furley, for putting up with me whilst all this went on. Thanks also go to my supervisors, Dr. Helen Pain and Dr. John Lee for their unfailing enthusiasm, support, feedback and encouragement; and also their determined efforts to help keep me employed and able to continue with this work. To my colleague Shari Trewin, I owe a singular debt of gratitude: for cheerily giving up her time in order to provide thoughtful feedback on often incoherent drafts of my work; for listening patiently to my half-baked ideas, musings and rants; and for opening my eyes to the world of Teviot chups!

My work has been influenced by discussions with and feedback from many others, particularly Rick Cooper and Ulrike Hahn, whilst latterly Lera Boroditsky has emailed abuse, stimuli and good ideas in equal measure, for which I'm sure I'm grateful. Many members of the analogy community have given freely of their time to answer my questions and comment on this work; I am particularly grateful to Ken Forbus and Dedre Gentner for the time they have taken in this respect. And I must also thank all my colleagues in the A.I. and Education Group, who have provided an extremely stimulating, supportive and friendly atmosphere in which to work. A special mention must also go here to Robin James, who made interesting and obscure suggestions regarding reading, and to Martin Corley, Tracy Packiam, Kaska Porayska-Pomsta, Rafael Morales, and Eshaa Alkhalifa, for cognitive coffee breaks and sanity checks; and to Judith Good, who turrned it all into a race.

My heartfelt gratitude goes to the innumerable volunteers who took part in my empirical studies, in some cases expending significant mental effort. In relation to the organisation of these empirical studies, the following individuals were most helpful: Sarah Darrington, Toby Lingstadt and Neil Meikle.

I am also grateful to Richard White, who first interested me in analogy, and to Dr. Paul Brna, who encouraged my earlier ideas. And I also would like to pay tribute to the community that was the Department of Artificial Intelligence at Edinburgh, for creating a friendly, challenging and truly interdisciplinary atmosphere; I feel it was a privilege to have belonged.

Finally, I'm grateful to Mark Keane and Usha Goswami for agreeing to actually read this stuff.

# Chapter 1

# Introduction

*'We talk of process and states, and leave their nature undecided. Sometime perhaps we will know more about them - we think. But that is just what commits us to a particular way of looking at the matter.'*
(Wittgenstein 1953, p102).

## 1.1 Modelling Concepts

Cognitive science has a strong interest in concepts, both literal and metaphoric: accounting for how the 'stuff of experience' is represented, manipulated and combined in the mind is a central concern of the field. The central aim of this thesis is to provide the groundwork for the investigation and modelling of human conceptual skills. Since in cognitive science conceptualisation is usually equated with categorisation, I shall acknowledge at the outset that no specific, complete model of categorisation will be presented. Indeed, although the process of categorisation is central to the concerns of this thesis, much of the material presented will not obviously address categorisation at all. Instead, much of the focus of this thesis will be on an 'extra-categorical' cognitive process: analogy.

The path chosen is not as quixotic as it might at first appear. A key tenet of science is that one should go beyond appearances, and seek out the causes or structures that underlie phenomena. And, as the remark from Wittgenstein at the head of this chapter seeks to show us, there is a pattern to the way in which problems regarding the understanding of mental processes[1] arise. The most important step in the production of such problems is the one that usually goes unnoticed: the deciding that one

[1] Wittgenstein actually considers the way *philosophical* problems concerning mental processes arise, though his analysis seems equally applicable to psychological problems.

manifestation of cognitive ability results from one process, and that another, apparently separate ability results from another. Wittgenstein's remark is a caution against the way that such assumptions can commit one to a particular way of looking at a matter, and also the way that the adoption of particular questions will - necessarily - preclude certain answers. What Wittgenstein seeks to warn us against is, in effect, becoming slaves to our own assumptions.

The contention at the heart of this thesis is that investigations of human conceptual abilities have relied too much on appearances. Because it appears, intuitively, that analogy and metaphor on one hand, and categorisation on the other, are different things, researchers have tended to treat them as manifestly different cognitive abilities. These assumptions, misplaced or otherwise, have not precluded fruitful progress in the quest to understand our conceptual abilities. Indeed, I shall argue that there is an extent to which these assumptions have helped research into analogy. However, in the end, it may turn out that what our assumptions of separateness preclude are just those answers that we are seeking.

"Extra-categorical" cognitive processes, such as analogy and metaphor, tend to be defined in relation to "straightforward categorisation." Yet despite the intimate linkage between analogy and metaphor on the one hand, and categorisation on the other (an intimacy manifest in these definitions), because at face value they appear to be so different from one another, the trend in cognitive science has been to treat analogy and categorisation as separable, distinctive phenomena that can be studied in isolation from one another.

This strategy has proved remarkably effective when it comes to the cognitive modelling of extra-categorical associations. A number of compelling and detailed models of the analogy process exist, and there is widespread agreement amongst researchers studying analogy as to what the key cognitive processes that determine analogies are. However, these models of analogy tend to assume some kind of fully specified category processing module which governs and determines ordinary, straightforward conceptual mappings. Indeed, this assumption is required in order to talk about analogy and metaphor in the first place: few theorists actually define analogy and metaphor *per se*, but all agree that analogical and metaphoric judgements can be defined in contrast to ordinary categorisation judgements.

Whilst analogy has proven to be a cognitive science success story (Forbus, Gentner, Markman and Fergusson, 1998, argue that analogy is one of *the* cognitive science success stories), the same cannot be said for categorisation research. The history of categorisation modelling is one of false dawns and set backs. At present, there is much agreement amongst researchers about how categorisation *shouldn't* be modelled, and none about how it should.

In this thesis, I shall argue that these differing histories are not unrelated; rather, there is a deep connection between the two. I shall argue that appearances have enabled analogy research to benefit from ignoring the problems involved in providing a 'straight' account of categorisation, whereas research into 'straight' categorisation has been unable to ignore the endless flexibility and plasticity of human conceptual abilities, of which analogy and metaphor are prime examples. The contention of this thesis is that analogy research has gone as far as it can on the unjustified - and I shall argue, unjustifiable - assumption that it is a distinct, separable process from categorisation. I shall argue that this distinction is based upon appearances, and that it cannot withstand detailed scientific scrutiny. Further, I shall argue that removing this distinction can allow progress made in studying analogy in isolation to inform the real question: that of explaining the processes that underpin all human conceptualisation, analogy, metaphor, and categorisation.

## 1.2 The Approach Taken

### 1.2.1 Cognitive Models

The central concern of this thesis is, then, mental representation, and the processes that interact with mental representations in human conceptualisation. The goal of research in this area is a working cognitive model of these conceptual process, thus it seems reasonable, at the outset, to outline a theoretical framework in which such a model can be situated and evaluated.

To take the problem of existing theories first: given that the theories and models discussed in this thesis are strongly computational, it follows from this that a key methodological component of the modelling process is that ideally any theory produced should be in some way amenable to computer implementation. The desire of researchers in this area is to articulate a psychologically plausible cognitive theory of conceptualisation that can be implemented as a model within a computer, and to show that this resultant model adequately demonstrates the theory.

As Johnson-Laird (1983) notes in respect to this ambition, like clocks, cognitive models need neither be wholly accurate nor correspond completely with the phenomena they seek to model to be useful. What is important, at the level of description at which they are proposed, is that models capture, in an appropriate way, the essence of whatever particular aspect of conceptualisation they seek to model. That is, the model should possess a 'relational structure' that corresponds to that of the particular aspect of the phenomenon modelled - at a similar level of description - such that the behaviour of humans carrying out that aspect of cognition is appropriately mimicked by the model. So, for instance, it might be that if we are modelling a particular aspect of conceptualisation, what is important is that our model capture the process that underpins the part of conceptualisation we are interested in, rather than being some kind of replication of conceptualisation per se. Models will tend to only embody essential or interesting features of whatever it is they are meant to represent.

As Goodman (1976) notes,

> "scientists and philosophers... have been forced to fret at some length about the nature and function of models. Few terms are used in popular scientific discourse more promiscuously than 'model'. A model is something to be admired or emulated, a pattern, a case in point, a prototype, a specimen, a mock up, a mathematical description - almost anything from a naked blonde to a quadratic equation - and may bear to what it models almost any relation of symbolization."
>
> Goodman (1976, p171)

There is a gap between our ordinary and our pre-theoretical treatment of 'models' and 'modelling.' My claim here, which owes much to Johnson-Laird (1983) is that in trying to understand a given phenomenon, humans construct some kind of internal model of it, and that their theories are attempts to express - in language - these 'theoretical models'.

One problem with this introspective, internalistic method of understanding is that many of the workings of our theoretical models may rely on assumptions or intuitions that we cannot fully externalise or explain (this point is especially acute in theorising about conceptualisation, as will be discussed further in Chapter 3). In the following review, a functional, rather pragmatic approach to this problem will be adopted: to at least close this 'intuition-gap' it will be assumed that theories ought to be describable in terms of an *effective procedure*, i.e. they should be described in terms of a step-by-step process that could be implemented in (and carried out by) some processor. If a theory is so described, then any reliance on intuitionistic or magical steps can be exposed or avoided; intuitive, magical steps will not be amenable to being performed by the processor as prescribed.

### 1.2.2 Cognition And Computation

Effective procedures, then, should specify how some function is to be carried out at an appropriate level of description. This in turn relies on a key assumption underlying the models described here (and indeed, of cognitive science in general), namely that the mind, and by extension, cognition, is in some way computational in nature; that cognitive functions are computational functions.

As Lee (1987) observes, this use of *computation* is not intended to be seen as some useful model of cognition (as if there were some helpful analogy to be drawn between the two), but rather the general claim in cognitive science is that "cognition literally is computation - in some sense" (Lee, 1987, p. 35). On this view, mental activity is (and mental functions are) regarded as computational, as the execution of algorithms. Thus the construction of cognitive models (which attempt to illustrate or explain some aspect of cognitive activity) will in effect entail the specification of an appropriate algorithm or program (I assume here that an algorithm provides an effective procedure for a computational function). In keeping with the majority of work discussed herein, this is the approach that has been adopted in this thesis.

### 1.2.3 Computation And Cognition

There are a number of methods for evaluating models proposed within this paradigm. In the evaluative framework described by Keane, Ledgeway and Duff (1994), separate evaluative levels are advocated for the evaluation of theories and models. Firstly, the '*informational constraints*' proposed by a theory are distinguished. Effectively, these are the theory's functions. In, for example, analogy, this corresponds to the level at which analogy is characterised, describing what needs to be computed to produce the appropriate output given certain inputs, i.e. such and such a correspondence should be found in, say, the analogy between the solar-system and an atom.

These informational constraints are considered separately from constraints at the "*algorithmic*" or "*behavioural*" level, which are specifically concerned with comparing the performance of given algorithms that implement informational constraints. This framework holds that algorithms can be compared with one another, or with the observed behaviour of people in a given task, in order to reduce the number of potential algorithms for a given function (or set of functions); i.e. behaviour constraints can be used to reduce the set of possible programs that could be produced to satisfy the informational constraints.

An example of a behavioural constraint that one might use to evaluate an algorithm for analogy is the consideration of working memory limitations. According to behavioural constraints, human working memory can easily become overloaded. Thus algorithms that reduce - or do not at least propose excessive - processing load should be preferred over those that do propose excessive processing load, even though the latter might satisfy informational constraints as well as the former.

Whilst the adoption of considerations such as behavioural constraints can provide a pragmatic method for evaluating candidate algorithms, it should be borne in mind that the statement that 'cognition is computation' does not entail a direct *identity* relation between natural and artificial computation. Identity statements regarding the way that an algorithm is implemented in the mind versus the way the way it can be coded in a computer are invariably speculative, owing to massive differences in the media of implementation. Moreover, the need for some caution here is best illustrated by the fact that even *behavioural* constraints can only be expressed in *informational* terms. For example, Keane, Ledgeway and Duff (1994) specify working-memory constraints and the influence of background knowledge as behavioural constraints on analogical mapping models. However, they can only specify the criteria for these constraints in informational - functional - terms.

This is a standard problem in theorising about mental processes. There is very little established theoretical bed-rock on which further theories can be based. The very nature of cognition seems to result in a scenario where the algorithms for implementing one set of informational constraints are evaluated against another set of informational constraints, whose algorithms remain, in turn, unspecified and unanalysed.

I do not, however, wish to argue here that the consideration of behavioural constraints is not worthwhile. It may be right that when a number of algorithms are specified for the same process, those with fewer processing demands should be preferred to those which make a massively heavy load on a processor. However such considerations seem ultimately to depend a lot on intuition. Moreover, in cognitive modelling, it is rare that theories tally sufficiently at the informational level to enable a final differentiation at the behavioural level. Given this, and the inherently functional nature of the paradigm - cognitive psychology - that provides the data for the specification of cognitive models, it would appear that the most important level of comparison between theories is at the informational level. What do the theories propose, and how well does this fit the functional requirements of the evidence?

The approach adopted herein is that it is only once theories and models have been analysed at this level, and it is clear that two or more proposals are informationally equivalent, that the further levels of analysis proposed by Keane, Ledgeway and Duff (1994) should come into play.

Thus the approach adopted here is that ultimately, theories and models must be judged on the adequacy of the explanations they provide, at the level of description appropriate to the phenomenon under consideration. According to this view, a good theory and its corresponding model should:

- lay down explicit constraints on the scope of a given phenomenon;
- detail the interactions of the mental process described with other processes;
- define its operations; and
- be consistent with the available empirical data.

It is these basic criteria that will form the basis for much of the evaluation in this thesis.

## 1.3 Thesis Outline

As stated above, this thesis begins with a consideration of analogy (Chapter 2), since analogy is the most successfully modelled of the areas of conceptualisation explored in this thesis. Chapter 2 examines the way analogy has been characterised in cognitive science research, and then details the two main theories of analogical recall and mapping in the literature. The workings of these models, and the main differences between them are laid out and explored. A consensus emerges from this review concerning the crucial role that similarities at a structural level between match items play in analogy.

The chapter then attempts to make explicit the critical conceptual assumptions that underpin all theories of analogy (and metaphor). Chapter 2 shows how analogy is widely defined in contrast to straightforward conceptual categorisation, and how all models of analogy rely upon reference to externally defined conceptual categories in their operations.

Chapters 3 and 4 explore theoretically these basic assumptions about the distinction between analogy and categorisation. In Chapter 3, I present a rigorous re-analysis of the theory of concepts attributed to Wittgenstein (1953). Wittgenstein's philosophical treatment of concepts has long been held to be foundational in the field, even though no detailed account of his actual views has been hitherto available.

The theoretical picture of concepts that emerges from a detailed reading of Wittgenstein's arguments is much at variance with the view of concepts and categories that has held sway in cognitive science research over the past 30 or so years. Chapter 4 presents a detailed review of the results of this research, and shows that despite this variance, Wittgenstein's account is compatible and consistent with these results. Furthermore, this chapter shows how the theoretical framework established by Wittgenstein can provide the theoretical glue required to make sense of the mass of data that has emerged from work in this area.

A particular picture of concepts and categories emerges from Chapters 3 and 4. This picture seems incompatible with the kind of account needed to validate a contrast definition of analogy. However, Chapter 2 shows that such a picture is essential to current characterisations of analogy in cognitive science. Furthermore, the review in

Chapter 4 shows how similarities at a structural level seem to play a key role in categorisation. In Chapter 5, I explore empirically the idea that seems to follow from this: that from the point of view of cognitive processes, the distinction between categorisation and analogy may be an artificial one. Three studies are presented in Chapter 5 which in various ways probe the distinction by presenting participants with materials usually used in analogy research to carry out categorisation tasks. The results of these experiments provide some support for the 'no distinction' hypothesis.

Chapter 6 presents a re-evaluation of the theories of analogy examined in Chapter 2 in the light of the evidence from Chapters 3, 4 and 5. Particular attention is paid to the analogical theories' external appeals to conceptual categories: if analogy and categorisation are the same process, then can such appeals be made? Two studies are presented that explore this question, and their results provide further support for a 'one process' account of analogy and categorisation. These results appear to show that in analogy, semantics can be mapped by use of a further, parallel analogical process, rather then by a substantively different process.

Finally, Chapter 7 discusses the research presented, summarises what has been achieved, and attempts to illustrate both the limitations of the presented work, and future avenues that should be explored. The implications of the 'no distinction' hypothesis for models of human conceptualisation are discussed.

# Chapter 2

# Cognitive Models of Analogy

*"There is no word which is used more loosely, or in a greater variety of senses, than Analogy" (Mill, 1882, p. 393)*

## 2.1 Introduction

In this chapter I present a review of the two foremost cognitive theories (and models) of analogy, Gentner's 'Structure Mapping' theory (Gentner 1983; 1989) and Holyoak and Thagard's 'Multi-Constraint' theory (Holyoak and Thagard, 1995). Many other theories and models of analogy exist in the literature (e.g. Indurkhya, 1987; 1992; Keane, Ledgeway and Duff, 1994; Hofstadter, 1995), however, the two theories that are the expository focus of this review are the most thoroughly developed, widely disseminated and closely related to a corresponding programme of empirical cognitive investigation of those that have been proposed.

To ensure that these two theory and model pairings can be clearly contrasted and analysed, some sacrifice has had to be made concerning the breadth of this survey in relation to its depth. This sacrifice has been made necessary as much of the later work described in this thesis is based upon the findings of studies reviewed in this chapter, and it has been necessary, therefore, to describe these studies in some detail. However, in order to best offset any possible deficiencies that this lack of breadth may result in, other theories will be discussed and described where they further illustrate points of contrast between Gentner's 'Structure Mapping' theory and Holyoak and Thagard's 'Multi-Constraint' theory.

## 2.2 What Is Analogy?

Analogy is much discussed and analysed in the cognitive science literature; what is meant by *analogy* is, however, rarely - if ever - stated. For example, Hoffman (1995), despite his stated aim to 'deconstruct, and then reconstruct, the concept of analogy' (p. 11), never actually offers a definition of what he means by 'analogy'. In this section, I shall review some of the characterisations of analogy given in the literature, and attempt to form a working definition of analogy.

Typically, characterisations of analogy are couched in terms of inferences and domains: Spellman and Holyoak (1996, p. 308) say, "analogical reasoning typically involves using a comparatively well-understood *source* domain as the basis for drawing inferences about a less well-understood *target* domain." Markman (1997, p. 373) largely concurs, describing analogy as "a powerful cognitive mechanism that permits two domains to be seen as similar on the basis of connected systems of relations that hold between them." Similarly, Clement and Gentner (1991, p.89) argue, "in an analogy, a familiar domain is used to understand a novel domain in order to highlight important similarities between the domains, or to predict new features of the novel domain."

All of which rather begs the question of how one defines domain, and the criteria one uses to decide whether mappings are inter- or intra- domain, since by these lights these are clearly the key factors that determine whether something is an analogy or not.

Other characterisations of analogy try to be more specific on the point. Gentner and Gentner (1983) offer a more structured and specific characterisation of analogy. They argue that an analogical statement such as:

1. The hydrogen atom is like the solar system

clearly doesn't indicate that all of a person's knowledge about the solar system should be transferred and attributed to the atom. Rather, it indicates that only some aspects of a person's knowledge should be transferred. Gentner and Gentner observe that whilst this may suggest that analogy is some kind of 'weak similarity' statement, where only some characteristics are transferred, such a characterisation would fail to distinguish between an analogical statement like (1) and literally a similar statement, such as:

2. There is a system in the Andromeda nebula that is like our solar system.

Gentner and Gentner argue that what is important in distinguishing between analogical and literal similarity statements is that whilst both convey similarities between two elements of comparison, the literal similarity statement (2) embodies an intention to convey similarities between both object and relations in the comparitors, whereas in the analogical statement (1) there is no intention to convey similarity between objects - e.g. the atom and the sun - but there is an intention to convey similarities between relations (e.g. orbital and perhaps other relational information).

Thus Gentner and Gentner distinguish analogical statements from literal similarity statements by distinguishing between the kinds of similarity the two propose. Taken in tandem with less precise comments about domains elsewhere, it seems reasonable to say that the overall picture of the concept of analogy that emerges from the literature is best summarised by Holyoak and Thagard (1995, p217) who note that analogy and metaphor "always connects two domains in a way that goes beyond our normal category structure" (pp 217): literal similarity - or categorical - statements propose similarities at object and relational levels, whereas analogies propose similarities at the relational level only. In the light of this, a reasonable working definition might be (to paraphrase Markman, 1997):

> Analogy permits two concepts to be seen as similar on the basis of connected systems of relations that hold between them.

## 2.3 Similarity - Based Retrieval

The focus of this thesis is upon analogical reasoning - in particular, analogical mapping - and its relationship to categorisation. Similarity-based retrieval - the process by which similarity-based remindings happen - has been included within its scope because:

- the processes are intimately related: it is in the very nature of two analogous items that they share some kind of similarity; I shall argue that analogy is a subset of the broader area of similarity-based transfer;
- especially in the area of explaining how an 'analogous representation' is accessed from memory in order, say, to further some explanation, the two problems have been treated as overlapping in the literature.

However, it should be noted that this thesis deals with similarity based transfer only in so far as treatments of the subject directly illuminate theories and models of analogical reasoning.

## 2.4 Rule Induction

Some treatments of analogy in the literature have tended to confine themselves to the very simple problem of the "*A* is to *B* is to *C* as *L* is to *M* is to *?*" variety: the induction of a single rule from one set of cases and the application of it in another (often termed *derivational analogy*: Sternberg, 1977; Rumelhart and Abrahamson, 1973; Mitchell and Hofstadter, 1989; Mitchell, 1993, Hofstadter, 1995). These examples, by limiting mapping to one clear relation, give little insight into how people interpret complicated explanatory analogies, which yield a number of potential mappings. I shall argue later in this thesis that this form of analogy is a particularly poor basis for the study of analogical mechanisms. This applies particularly to the central concern of this thesis: the relationship between analogy and categorisation in cognition.

Derivational analogies make especially heavy assumptions about the content and nature of the conceptual representations that underpin individual elements within them: in particular, about the way *A*, *B*, *C*, *L* and *M* must be represented in order to infer *anything* about *?* in "*A* is to *B* is to *C* as *L* is to *M* is to *?*".

Models resulting from such theories have little or nothing principled to say about the kinds of constraints that act upon the selection of any particular set[1] of mappings which leads to their accounts of analogy being seriously incomplete (this question will be further discussed and illustrated later in this thesis).

Since this study is concerned with high-level analogical mechanisms - the imprecise, but meaningful relationships that can exist between complex structures - rather than

[1] Hoffman (1995) refers to the "semantic flexibility" of such analogies. This semantic flexibility results in their representations being dependent on what their constituent words are interpreted to mean. And these meanings are massively unconstrained in derivational analogies, with words occurring individually, in no particular context. Take for example fridge : food :: pocket : ?? - if we assume the hidden rule is 'contains', then ?? might be anything one might plausibly find in a pocket, such as 'handkerchief'. However, suppose we remember the idiomatic expression 'cold cash'. We might then assume ?? is cash, and the rule is 'keeps cool', and hence reject 'handkerchief' (see Hoffman, 1995, for many more illustrations of the endless ambiguities inherent in derivational analogies).

simple low-level rule inductions, this review will deal only with those theories that attempt to provide explanations for high-level analogies. However, it should be noted that whilst models of high-level analogical reasoning can also be applied to rule inductions (Burns and Holyoak, 1994), existing models of rule-inductions have not been shown to be capable of simulating high-level analogical reasoning.

## 2.5 Analogical Reasoning

It is widely accepted that analogy is a central cognitive process. The use of existing information in order to explain novel concepts or generate new ideas occurs with great frequency over many levels of cognitive activity, from the metaphors of everyday speech to the classification of visual perceptions (in a classic 'Peanuts' cartoon Charlie Brown asks Schroeder what he can see in the clouds - Schroeder's rather sophisticated response - "Bucephalus?" - causes Charlie Brown to reject mentally the ideas he had formulated, 'ducky' and 'horsey').

The problem of analogy can be broken down into a number of basic sub-processes :

- *Accessing*: i.e. the problem, mentioned above, of explaining how an analogy is accessed from memory. How is it that the stored understanding of two analogous processes in memory- say the workings of the solar system and the hydrogen atom - enable us to select one as illustrative of the other?
- *Mapping*: how does the person to whom the analog has been presented map their prior knowledge of the solar system onto their, perhaps somewhat hazy, knowledge of the hydrogen atom in order to gain a firmer understanding? What common properties must be mapped between two domains for them to be considered analogous?
- *Inference*: following on, what properties and relationships in the solar system can be inferred as also pertaining to the atom? Which inferences are valid, and which invalid, and why?

Other sub-processes can include: the representation of initial start states or problems - the way that an individual represents a problem can affect the success or otherwise of subsequent analogising (Clement, 1988; Novick, 1988); adaptation of base analogies (changing the problem representation to explore different analogies, Keane, 1994); subsequent learning (including schema induction, Gick and Holyoak, 1983); and verification. The latter, *verification* is often considered to be a separate sub-process of analogy. However, since it would appear that some element of verification is an essential part of *any* psychological sub-process it will be considered wherever appropriate. Moreover, it would appear that verification - as an analogical process - can easily be given too much prominence: an over-concentration on what counts as the right analogical transfer can lead to a tendency to demand a spurious kind of precision: *foo* is an analog of *bah*, and any model must always make the right transference from *foo* to *bah*, regardless of how *foo* and *bah* are represented. Analogical mapping is necessarily contingent upon representations, and hence any description of analogy is a description of a *contingent* process. Losing sight of this, and demanding a spurious precision from a psychological theory or a computational model, can lead to features that are inherent within the analogical process being seen as problems when they appear in models and explanations of it. The relative imprecision of the analogical reasoning process is an important aspect of its nature, and any explanation of the process of analogy must capture the nature of these problems.

In this review chapter I shall concentrate on two of these analogical sub-processes, mapping and retrieval. These have received most attention in the literature, and are generally seen as the core of the analogy process. Moreover, as I hope to show later in this thesis, the specification of these processes, and especially that of mapping, has wide reaching implications for the way analogy, and hence other proposed sub-processes, are seen in relation to wider cognitive processes.

## 2.6 Analogical Mapping

### 2.6.1 Gentner's Structure Mapping Theory

Gentner (Gentner, 1983; Gentner and Gentner, 1983) proposed the Structure Mapping Theory as an attempt to explain how it is that two domains can be considered analogous, and in particular how it is that correspondences between analogs[2] from two domains can be mapped.

---

[2] For the current discussion, I shall use "analogs" to refer to items from separate domains that are perceived to be analogous. The theoretical basis for this 'definition', and its sustainability in the light of a theoretical and empirical evaluation, will be considered in chapters 3 and 4.

Structure mapping proposes that the mapping and inference between two domains can be achieved by assigning correspondences between objects and attributes and then mapping predicates with identical names. In order to do this, Gentner assumes a predicate like representation (figure 2.1), distinguishing between *objects, object-attributes* and *relations.*

- Object-attributes are those predicates that have one argument and describe object properties, e.g. RED(lobster).
- Relations are divided into a hierarchy of orders. Relations between objects, e.g. UPSETS(stomach,lobster) form the lowest order of relations. Predicates which describe different levels of relationships between relations e.g. CAUSE( UPSETS(stomach,lobster), DRINKS(alka-seltzer,diner) ) then form the higher orders, with each higher order of predication implying a greater embedddedness of the lower order predicates it comprises.

---

**Solar System**

YELLOW(sun)
MASSIVE (sun)
HOT(sun)
ATTRACTS(sun,planet)
MORE_MASSIVE_THAN(sun,planet)
REVOLVES_AROUND(planet,sun)
CAUSE(ATTRACTS(sun,planet), REVOLVES_AROUND(planet,sun))
HOTTER_THAN(sun,planet)

**Hydrogen Atom**

ATTRACTS(nucleus,electron)
MORE_MASSIVE_THAN(nucleus,electron)
REVOLVES_AROUND(electron,nucleus)
CAUSE(ATTRACTS(nucleus,electron), REVOLVES_AROUND(electron,nucleus)

---

**Figure 2.1**: Predicate representations of the solar system and a hydrogen atom

The theory itself comprises two parts: *mapping rules*, and the *systematicity principle.*

Mapping rules state that:

- attributes of objects are not mapped
- relations between objects are preserved

Whilst the systematicity principle requires that:

- complex higher order relations (e.g. CAUSE above) are mapped preferentially followed by relations that constitute the higher order arguments.

This latter theoretical constraint is intended to capture the notion that analogy conveys a system of connected knowledge, rather than an assortment of independent facts:

> "structure mapping stems in part from the observation that useful analogies, such as those used in science or education, involve rich, interconstraining systems of mappings between two domains, rather than a set of independent correspondences"
>
> (Clement and Gentner, 1991, pp 91-2)

Finally, as a background assumption, the theory embraces the idea of 'canonical conceptual representation.' As a result of this assumption, only predicates with identical names are mapped onto one another, with the assumption that non-identical but conceptually similar predicates will be matched elsewhere according to established conceptual higherarchies. Gentner et al (Gentner, Ratterman and Forbus, 1993; Forbus, Gentner and Law, 1995) describe the canonical conceptual representation assumption thus:

> 'SME's constraint of matching identical predicates assumes canonical conceptual representations, not lexical strings. Two concepts that are similar but not identical (such as "bestow" and "bequeath") are assumed to be decomposed into a canonical representation language so that their similarity is expressed as a partial identity (... "give")'
>
> (Gentner Ratterman and Forbus, 1993).

In practice, this identicality constraint is effectively a third mapping rule: structure mapping theory insists that mappings only occur between predicates of the same name. Thus the question of whether, say ORBITS and REVOLVES_AROUND are matching predicates is settled *outside* of the theoretical parameters of structure mapping theory.

Gentner *et al* assume that the match between ORBITS and REVOLVES_AROUND is settled by these canonical conceptual representations, and thus the mechanisms by which they are matched can be assumed - i.e. the existence of an appropriate mechanism is taken as a background assumption - by both their theory and model.

The psychological plausibility of the structure mapping theory was tested in Gentner and Gentner (1983) by comparing the performance of test participants in the domain of simple electric circuits when taught using one of two different analogies.

The first of these was the analogy between electricity and circuits and water and hydraulic systems. It was posited that this would best enable students to predict the behaviour of circuits with two batteries in series or parallel, rather than those with varying resistor combinations, since the relationship between batteries and hydraulic reservoirs has more correspondence than that between water flow through restricted pipes and electrical resistors.

The second analogy compared electricity to crowds of people moving through passages. In this case, the correspondence between resistors and narrow passageways is strong, whereas there is less correspondence in relating a battery voltage by how hard people push, thus it was expected that this analogy would present problems in enabling the prediction of the consequences of battery combination.

Gentner and Gentner's experiments backed up the structure mapping hypotheses - a higher proportion of correct answers to battery combination questions were observed in participants taught using the hydraulic model, and a corresponding higher proportion of correct answers to resistor combination questions were obtained by participants taught using the teeming crowds analogy.

On the whole, these early experiments lent some credence to the notion that the structure of a base analogy can affect the structure of a corresponding mental model when used in teaching, though these experiments were not sufficiently sensitive to really test the principle underlying the structure mapping theory.

One drawback of these early experiments is that they are open to the criticism that they use too arbitrary a representation (Carrol and Mack, 1985). The primitives used in the examples are chosen selectively, and the inclusion of a larger range of domain relations would result in the matches appearing far less sound than they do. For example, it is claimed that Gentner can give no account as to why HOTTER_THAN in figure 2.1 should not be mapped onto the atom representation. Holyoak (1985) argues that expanding the representation in order to utilise the systematicity principle will be of little help, since on another expansion the relation HOTTER_THAN will form part of a higher order system (involving the warming of planets, conditions for the existence of life, etc.). However, as Gentner (1997) observes, this argument largely misses the point: structure mapping explains why HOTTER_THAN is not mapped given *these* representations. It might well be that provided with representations expanded in the way that Holyoak proposes, HOTTER_THAN would be mapped, and different inferences would be supported - that would all depend on *those* representations, and what mappings resulted from their comparison. Given the intimate relationship between analogy and the way an intended analog is represented, this is not surprising.

Holyoak claims that his general objections are indicative of a serious problem. If a theory (or model) uses no goal or domain information, and relies purely on structural inferences to explain analogy, he claims it will have trouble when, given a number of networks of higher-order relations between domains (i.e. several causal networks), giving an account as to why one particular set of mappings was selected for, and considered relevant to, making the analogy.

### 2.6.2 Holyoak And Thagard's 'Multi-Constraint' Theory Of Analogical Transfer

Holyoak and Thagard's theory (Holyoak, 1985; Holyoak and Thagard, 1989; Thagard *et al*, 1990; Holyoak and Thagard, 1996) attempts to bring these broader considerations into a theory of analogy (though it should be noted that it is concerned primarily with problem solving). To some extent this theory has been developed in competition with Gentner's structure mapping theory. It attempts to capture the intuition held by some researchers that goals play some part in the analogical process, particularly in mapping (Gentner's theory precludes any such influence). The scope and ambition of this theory has changed considerably over time: what is presented here is a sketch of its latter incarnation.

Holyoak advocates an explanation of analogy in terms of a goal-driven processing system, and that mappings be controlled by the system's goals:

> "Within the pragmatic framework, the structure of analogy is closely tied to the mechanisms by which analogies are actually used by the cognitive system to achieve its goals." (Holyoak, 1985, p76).

Holyoak and Thagard see the analogical mapping problem as one of explaining how the large number of possible mappings between domains can be evaluated and a subset of these used for the transfer of information between domains.

They suggest that this subset emerges from an attempt to balance the different influences upon the mapping process. More specifically they regard the process as an attempt to simultaneously satisfy several constraints (described in full below). The first group of these are structural, and have strong correspondences to the Structure Mapping model:

- logical compatibility
- role identity
- uniqueness
- relational consistency

However Holyoak and Thagard also consider the following constraints integral to the mapping process:

- pragmatic usefulness
- semantic similarity

#### 2.6.2.1 Structural Constraints

*Logical compatibility* ensures that mappings are only considered if they are between entities of the same "type". Thus, in the solar-system / hydrogen atom analogy the predicate MORE_MASSIVE_THAN cannot be matched with the object electron. Similarly, a mapping between the predicates HOT and REVOLVES_AROUND will not be considered as they take different numbers of arguments. This primarily syntactic constraint is intended to ensure that mappings between different levels of description are not attempted. For example single-argument predicates, such as HOT, tend to be purely descriptive, specifying a particular attribute of an object. Multi-argument predicates describe relationships between objects, and so can be considered to represent a higher level of description. The argument is that mappings between different levels of description are not going to be productive and this constraint serves to eliminate any potential mappings of this kind.

A further hurdle potential mappings must overcome in this model is the *role identity* constraint. This assumes that the base and target domains can be divided at a higher level of description than that at which the mapping takes place. In the use of analogy in problem solving, upon which the authors focus predominantly, this means the domains may be redescribed in terms of a start state, the problem goals, and the operators that can be used to try and achieve these goals. Role identity then limits mapping to relations and objects that appear in the same part of the domain definition. This provides a weak pragmatic influence in that elements can only be considered for mapping if they play a similar role in both domains. These constraints are intended to limit the number of potential mappings that are considered in the evaluation phase, which utilises the next set of constraints.

Holyoak and Thagard assume that each element in the base domain will map onto one and only one element in the target domain, and *vice versa*. Thus there will be competition between members of the set of potential mappings between one base element and a number of possible target elements, i.e. if HOTTER_THAN in the base maps onto HOTTER_THAN in the target then it cannot map onto LESS_MASSIVE_THAN in the target. Hence any factor which serves to increase the level of support for one particular mapping will consequently act to decrease support for all the others.

Relational consistency acts to ensure that any mappings between the base and target domains must be consistent. Thus if mappings between structural elements receive support, mappings between the structures themselves, and any other elements, are also supported. For example support for the mapping between sun and nucleus applies to the mapping between the higher order relations MORE_MASSIVE_THAN in the base and target as well; correspondingly the mapping between planet and electron will also gain support. This corresponds quite closely to Gentner's systematicity principle.

Two possible influences upon the mapping process rely on information other than that found in the basic domain representations. These are the pragmatic and semantic constraints:

### 2.6.2.2 Pragmatic Constraints

The importance of an element (object or relation), whether in the base or target domain, is another consideration in the mapping process. An element's importance is defined in terms of how useful the element is in satisfying the current goal (or subgoal) of the "analogiser". Thus any mappings involving these "useful" elements will receive more support than those involving less useful elements. When our example analogy is used to explain the relative motion of sub-atomic particles, mappings involving YELLOW and HOTTER_THAN are going to be less favourably considered than those involving REVOLVES_AROUND, since the former are not utilised in satisfying any explanatory goals. Pragmatic information can come from either the base or target domains. For example the information that an element is an important constraint on any potential problem solutions is likely to come from the target domain. Alternatively information regarding the contribution a particular element can make in actively producing a solution will come from the base domain in most cases.

A considerable body of evidence has been built up to back the idea that pragmatic factors have a role to play in analogical reasoning (Spellman and Holyoak, 1992; 1996). However, whilst there is evidence that goals can influence *the analogical reasoning*, any evidence that goals affect *analogical mapping* is far less clear. Spellman and Holyoak (1996) say of their findings with regards to pragmatic influences: "it is no doubt possible to develop an account of the present results (or any other pattern of mappings) in terms of some pre- or post- mapping process" (Spellman and Holyoak, 1996 p. 343). Such an architecture was described by Gentner (1989) allowing pragmatic constraints to influence analogical reasoning, but not analogical mapping. This architecture accommodates the idea that goals and plans will act to select certain aspects from a particular problem under consideration, and that these will be used to seed a search of long-term memory for a suitable analog. Once this is found, it can serve as a base for subsequent mapping with the original target problem. Any output mappings will then be evaluated against pragmatic goals and plans to gauge the success or otherwise of the analogy (with the process being iterated if necessary).

This architecture allows Gentner to let structure mapping theory reflect the pragmatic factors which have been shown to influence analogy (in terms of problem representation and output evaluation) whilst preserving a wholly structural account of mapping.

Keane (1988) makes the observation that an obvious lacuna here is the lack of a precise statement of the way pragmatic influences work within the proposed cognitive architecture. However, whilst this is indicative of the want for detail in Gentner's account and architecture, it is also a wider indicator of the absence of any statement of an effective procedure for the influence of pragmatics in theories that advocate their influence. The level of detail in the evidence for pragmatic factors collected so far does not provide any revealing or specific clues as to the *process* by which pragmatics influence analogies. Certainly, this evidence is of the wrong level of description to resolve any debate regarding the influence of pragmatics on analogical mapping. Again, as Spellman and Holyoak (1996) note of their findings with regards to pragmatic influences: "it is [...] possible to develop an account of the present results [...] in terms of some pre- or post- mapping process" (Spellman and Holyoak, 1996 p. 343), such as the model proposed by Gentner (1989). Despite this, Spellman and Holyoak make the claim that "the obtained pattern [from their study] can be explained parsimoniously in terms of an inhibitory mechanism for control of selective attention, which influences a process of mapping by constraint satisfaction" (Spellman and Holyoak, 1996 p. 343).

Clearly however, in so far as an effective procedure for the way pragmatics influence analogical mapping is still wanting, this debate is still very much open; and no such procedure is forthcoming in the literature describing pragmatic influences on analogy. The answer to the question of whether pragmatic factors play a part in analogical mapping seems to be determined at present by the intuitions of whoever is theorising about them - or at least by their intuitions regarding cognitive parsimony - as indeed are the details of how this part is played.

#### 2.6.2.3 Semantic Constraints

Holyoak and Thagard suggest that the most useful mappings are likely to come from elements which are semantically similar. In the solar-system / atomic structure analogy, predicates with identical names can be regarded as more similar than those with different names. In more complex examples the method of determining relative similarity is more difficult.

Holyoak and Thagard (1989) argue that a requirement of strict identity between corresponding relations - as proposed in structure mapping theory - is unsatisfactory, since people can readily find mapping between non-identical relations (Holyoak and Thagard 1989; Burstein, 1986). According to structure mapping theory, only predicates with identical names are mapped, thus if the planets of the solar-system were represented as SMALLER_THAN the sun, and electrons as LESS_MASS_THAN a nucleus, then structure mapping theory would not allow a mapping between the two relations. On the other hand, whilst multi-constraint theory specifies a preference for mapping identical relations, it can also capture less direct semantic similarities in order to model the semantic similarity between SMALLER_THAN and LESS_MASS_THAN.

Holyoak and Thagard argue that this shows a significant weakness in Gentner's theory:

> "with its emphasis on structure to the exclusion of all other constraints, [structure mapping theory] does not simply discourage mappings between non-identical but semantically similar items; it does not even permit them." (Holyoak and Thagard, 1995, p. 258)

Holyoak and Thagard are, however, vague as to what semantic similarity comprises. Nor, as with goals above, do they translate evidence that semantic similarity plays a part in analogical reasoning to an effective procedure for the role of semantics in analogical mapping. Thus, in multi-constraint theory, the semantic similarity constraint is regarded more as a heuristic than a firm rule, and can be applied in differing strengths at various stages of the mapping process.

Holyoak and Thagard make no claim as to any particular model of semantics. They merely observe that in representations of analogical comparitors, object similarity can "potentially be reduced to predicate similarity: two objects are similar to the extent that they serve as arguments of similar predicates. Predicate similarity may in turn be analysed in terms of feature overlap (Tversky, 1977)" (Holyoak and Thagard, 1989, p. 301).

#### 2.6.2.4 Discussion

Holyoak and Thagard (1989) reduce the logical compatibility and role identity constraints to restrictions on the building of the mapping network, and these restrictions are regarded as less rigid than the constraints of isomorphism (the uniqueness and relational consistency constraints), and general pragmatics (semantic similarity and pragmatic centrality). Thus multi-constraint theory largely involves a computation of structural similarities between two comparitors, since as far as pragmatic and semantic constraints are concerned, Holyoak and Thagard state, "for the present purposes [of developing multi-constraint theory into a cognitive model of analogy] it will simply be assumed that the mapping component can receive a numerical index of the degree of semantic similarity between two predicates, and of the pragmatic centrality of elements of the analogs" (Holyoak and Thagard, 1989, p. 306).

### 2.6.3 ACME - The Analogical Constraint Mapping Engine

ACME (Holyoak and Thagard, 1989) is a program which models this process as a constraint satisfaction network. This constraint satisfaction network is used to evaluate the potential mappings. It comprises nodes which represent mappings, and links which reflect potential support for these mappings. The logical compatibility and role identity constraints are used to determine the mapping nodes created, and then the uniqueness and relational consistency constraints are utilised in order to specify the links that are made between them.

Thus, in keeping with the logical compatibility constraint, mappings are limited to predicate - predicate matches, where the predicates are of the same arity, and object - object mappings. The role identity constraint is implemented by restricting mappings to those elements in the corresponding parts of the specification, limiting start conditions to maps with start conditions, and goals to maps with goals, etc.

Each node will exhibit a certain level of activation, the level of which is governed by the amount of support it is receiving at any given time from neighbouring nodes. Thus in addition to its mapping information, each node has an activation value which reflects this support, an output value (which is the same as the activation value when the former is positive, and zero when it is negative), and a change value, which reflects the difference between the node' s current activation value and its activation value on the previous cycle (this information is used in determining when the network has settled).

A node is connected to its neighbours by links, which are either excitatory in situations where the node is receiving positive support, or inhibitory where support is negative. The choice of whether links should provide positive or negative support is governed by the uniqueness and relational consistency constraints. Thus nodes for competing alternative mappings are connected by inhibitory links, whilst nodes which represent mappings of elements from the same structure would gain support (since nodes are bi-directional, such support will be mutual).

Each link has an associated weight, which can be varied in order to modify the amount of support it can provide. In the case of excitatory links, this weighting is positive, whilst for inhibitory links it is negative[3].

---

[3]To perform an evaluation, the activation values of the mapping nodes are set to a value of 0.1, and the output value of the semantic and pragmatic nodes is set to 1. The activation value of each node is

| Base | Target |
|---|---|
| T1: D(a,b) | S1: M(x,y) |
| | S2: N(x,z) |
| | S3: N(w,y) |

**Figure 2.2**: A Problem for the Early ACME (from Holyoak and Thagard, 1989, p 311).

However, in order to overcome some difficulties encountered using the earlier version Holyoak and Thagard (1989) added a new component, *propositional identifiers*. The mappings and links created in the above fashion encounter difficulties when they are used in circumstances like those illustrated in figure 2.2. If the evidence provides high activation to the object mapping units $a=x$ and $b=y$, then intuitively this should suffice to establish that predicate D maps to M. However, in ACME the D=N argument will receive more support because the D=N mapping node will receive support from both N($x$,z) and N(w,$y$), since final evaluation depends upon possible proposition correspondences.

To avoid this undesirable outcome in later versions of ACME (Holyoak and Thagard, 1989) propositions are labelled with identifiers, and proposition mapping-units have been introduced, which have excitatory connections to both the corresponding

---

then calculated according to the formula

new = current * (1-decay) + enet (max - current) + inet (current - min)

where

*new* = new activation value

*current* = current activation value

*decay* = constant defining the rate at which the activation of a node will decrease with no input

*enet* = sum of all excitory inputs on the node

*inet* = sum of all inhibitory inputs to the node

($net = \sum o_1 * w_1, o_2 * w_2, ...., o_n * w_n$, where $o$ = output and $w$ = weight on each link to the node)

*max* = maximum permitted activation value

*min* = minimum activation value

The calculations for each node are repeated until the network has settled (either because the change value of each node has fallen below a threshold, or a predetermined number of cycles have been performed).

predicate-mapping units and the object-mapping units. Thus, in the example in figure 2.2, links between the proposition identifiers T1 and $S_n$ are created, which results in the mappings T1=S2 and T1=S3 becoming competing mappings: thus the link between them will be inhibitory, with the result that T1=S1 will be the favoured mapping.

It is worth noting here that, regardless of the intentions of Holyoak and Thagard, the addition of propositional identifiers further strengthens the dominating influence of structure in the ACME mapping process (White and Brna, 1990).

### 2.6.4 Evaluating ACME

The first obvious criticism of Holyoak and Thagard's model is that it is limited in scope. It only applies to situations in which an analogy is used within the context of some pre-existing goal that provides the analogiser with an idea of the kinds of inferences and matches that need to be made. A common use for analogy is in learning and teaching, and in this context it is difficult to give a satisfactory account of how goals can influence particular mappings. Indeed, the selection problem re-emerges. If an analogy offers two potential sets of mappings and inferences, a goal based upon a desire to understand will not provide the analogiser with any grounds for choosing one set of mappings over another (this point will be further examined below).

A second cause for concern is the dominating role played by structural considerations in ACME. If Holyoak and Thagard are right in their assertions that pragmatic and semantic factors are necessary to analogical mapping, then the unsatisfactory implementation of these factors in ACME (an arbitrary weighting that has little real significance on the settling of the network, White and Brna, 1990) must undermine support for the system. Moreover, the inadequacy of Holyoak and Thagard's description of the way these factors influence the mapping process - in neither case are these constraints given a theoretical description that could be realised in terms of an effective procedure - provides little indication of how this implementation might be improved.

However, Holyoak and Thagard claim - and their simulation results confirm - that ACME has provided psychologically plausible results in the course of analogical experiments (Holyoak and Thagard, 1989; 1995). This leads to a quandary: whilst Holyoak and Thagard's theory stresses semantic and pragmatic factors, their implementation ultimately utilises structural information in determining mappings. If their theory is right, then their implemented model is inadequate; on the other hand, given that their model provides psychologically plausible output, the weight of evidence suggests that their theory over-stresses the importance of pragmatic and semantic factors.

This leads to a further problem. The inclusion of pragmatic and semantic factors in Holyoak and Thagard's model was partly prompted by a desire to avoid the problems encountered by Gentner's structure mapping theory: the over-reliance upon representation, and the fact that structural considerations alone give one no basis for choosing between two equally structurally viable interpretations of an analog. However, Holyoak and Thagard's model is ultimately structural. Its implementation, ACME - the model is the most formal statement of the theory - ultimately relies on structural factors in determining mappings. Thus it should be noted that to a large extent, these criticisms apply as equally to Holyoak and Thagard's model of analogical mapping as they do to Gentner's.

## 2.7 Relational Structure And The Selection Of Mappings

If analogy is characterised as being a process whereby a partial similarity match is made between certain common features of different domains, then there must be an account of the factors that influence the deciding of which commonalities are mapped and which are disregarded. To take the plumbing/electricity example, a number of relations could be mapped: both are distributed in networks from central supply companies (all of which are newly privatised natural monopolies...). Both can be used in cooking; degree of pressure determines flow rate, etc. What account can be given for people's observable fluency in performing the correct analogical mappings? This then is the selection problem: given that two domains might have a number of causal networks in common, each available for mapping, how is the appropriate analog structure selected?

Clement and Gentner (1991) sought to add further support to the claim that relational structure is, in fact, the key selection constraint in analogical mapping. The value of a given match depends not only upon the component match itself, but on other matches to which it is connected; the systematicity principle.

**Diner**

RED(lobster)
GOURMET(diner)
LIVES_UNDER(rock,lobster)
DRINKS(alka_selzer,diner)
UPSETS(stomach,lobster)
CAUSE( UPSETS(stomach,lobster), DRINKS(alka_selzer,diner) )

**Cricket fan**

LIVES_UNDER(threat_of_humiliating_defeat,cricket_team)
DRINKS(gallons_of_lager,cricket_fan)
EATS(greasy_burger,cricket_fan)
UPSETS(cricket_fan,humiliating_defeat)
CAUSE( UPSETS(cricket_fan,humiliating_defeat),
DRINKS(gallons_of_lager,cricket_fan) )[4]

**Figure 2.3**: More and less analogous elements within two representations

According to the theory (restated in Clement and Gentner, 1991), the attributes of the objects of the first analog in figure 2.3 are not mapped, whilst the relations UPSETS and DRINKS are mapped preferentially because they play a part in the higher order relation CAUSE. The support for the mapping of LIVES_UNDER receives less support (despite the fact that our cricket fan might wish his team to go live under a rock) since it does not play a part in any higher order relation.

Clement and Gentner argue that systematicity does in fact provide an account of selection. To back up this position, they point to studies such as that by Holyoak and Koh (1987), which varied the degree of structural correspondence between analogous problems and, consistent with the systematicity principle, they found that when the

[4] It is interesting to note that intuitively, the efficacy of this analogy does not appear to alter if DRINKS(alka_selzer,diner) is changed to EATS(alka_selzer,diner), or even perhaps SEEKS(alka_selzer,diner). This does, at surface level at least, seem to illustrate the divergence between semantic and structural factors in analogical mappings that Holyoak and Thagard argue for. Whether this insight should be carried through at a theoretical level will be discussed at greater length below.

causal networks describing the source and target problems differed, participants were less likely to transfer the solution from source to target.

| **Base: The Tams** | **Target: The Robots** | |
|---|---|---|
| | **Version 1** | **Version 2** |
| Consume minerals with underbellies | Gather data with probes | Gather data with probes |
| **Exhaust minerals in one spot and must relocate on the rock** | **Exhaust data in one place and must relocate on the planet** | Internal computers overheat when gather a lot of data |
| *So stops using underbelly* | *So stops using probes* | *So stops using probes* |
| Born with inefficient underbelly | Designed with delicate probes | Designed with inefficient probes |
| **Underbelly adapts and becomes specialised for one rock** | Robots cannot pack probes to survive flight to another planet | **Probes adapt and become specialised for one planet** |
| *So underbelly can't function on new rock* | *So probes can't function on new planet* | *So probes can't function on new planet* |

*Note. Key facts are shown in italics. Matching causal information is shown in boldface. In experiment 2, italicised facts were removed from the target.*

**Figure 2.4:** Relational structures of the base domain (the Tams) and the target domain (the Robots). From Clement and Gentner (1991). It should be noted that in the actual experiments, efforts were made to avoid the extensive surface similarities in sentence structure that are included here in order to aid comprehensibility.

This evidence suggests that the criticisms voiced by Carrol and Mack (1985) - that the elements used in representations in structure mapping theory are incorporated too selectively - might in fact apply equally to the way people perform analogical transference as to the way it is represented in the structure mapping model, and as such this might actually lend weight to the psychological plausibility of the theory, rather than detract from it. The evidence suggests that in order for people to perform analogies, it is important for them to be presented with a base that is properly represented. The solar-system / atom analogy would break down if someone using it to teach Rutherford's model of the atom chose to concentrate on inappropriate details in the base (the heat of the sun, the size of the planets, their colours, etc.) to the exclusion of relational details. If the appropriate relations were glossed over, or others emphasised, then the analogy would not work. Put simply, structure mapping theory utilises selective representations in explaining analogy because people use selective representations in performing analogies.

In order to further test the claim that systematicity can act as a selection filter during analogical transference, Clement and Gentner performed three experiments that looked separately at two of the components of analogical mapping:

- matching existing information in the base and target, and
- inferring new information about the target that follows from the analogy with the base domain.

The first tested whether systematicity constrains the matching process. A novel set of analogies were created. Each consisted of a base and a target passage describing objects or organisms on fictional planets. Each passage contained two main paragraphs, one describing a causal structure that matched between base and target, and one that did not match. Participants could choose which of two facts (lower-order relations) to map from base to target domain. In all cases, both facts were equally acceptable as independent matches. However, facts varied as to whether they were part of a shared causal network (relational structure). It was suggested that if systematicity plays a part in the matching process, then participants would show a corresponding preference for those matches that are embedded within a matching causal system rather than those where causal systems are unconnected.

The base of one analogy involved creatures called *Tams* who live on a distant planet. Encyclopaedia-like passages were given to participants, the essential elements of which were:

> Paragraph 1: The Tams live on rock and can grind and consume minerals from the rock through the constant action of their underbelly. However, periodically they run out of minerals in one spot on the rock and must relocate. At this time they stop using their underbellies.

> Paragraph 2: Although at birth the Tams have rather inefficient underbellies, eventually the underbellies adapt and develop a texture that is specially suited to the rock the Tam lives on. As a consequence, a grown Tam's underbelly cannot function on new rock.

So each paragraph has a key fact and a causal antecedent. The key facts in this domain are:

1. The Tams sometimes stop using their underbellies

2. The Tams' underbellies cannot function on new rock.

The analogous domain was called "The Robots", and described robots who used probes to gather data from planets. Figure 2.4 gives an outline of both domains. The left column shows the two causal structures of the base, with each key fact shown in italics, and the causal antecedents shown in boldface. The central column shows version 1 of the Robot domain which, like the Tam domain contains two causal structures. The key facts in the target match those in the base:

1. The {Tams, Robots} sometimes stop using their {underbellies, probes}

2. The {Tams', Robots'} {underbellies, probes} cannot function on new {rocks, planets};

however, although the key facts in the target match those in the base, only the first key fact is linked to a causal system that is matched by the base domain, which Clement and Gentner call a *shared-system* key fact. They predicted that participants would prefer this shared system fact in mapping to the other key fact, which does not share a causal antecedent.

The second and third experiments explored the notion of systematicity acting as a constraint upon inferences carried over from the base domain to the target. Two facts were present in the base domain that were equally plausible as inferences about the target, but only one fact was linked to a causal system shared by both the base and the target. It was expected that participants would not just select any base fact, but would rather infer a fact that follows from a shared causal network. Experiment 3 differed from experiment 2 in that the base story was presented before, but not during, the inferencing tasks, hence participants had to rely on their memory of the base in choosing their inferences.

Experimental results supported these predictions. In the first experiment, participants preferred the matching fact that was embedded within a matching causal system, whereas the control group, who saw only the target domain showed no preference for the shared system fact. Clement and Gentner claim that this is evidence that the systematicity principle does constrain matching; "analogical matching is not merely a feature-by-feature decision: Analogical matching concerns systems of predicates, not individual predicates". Experiment 2 also provided support for systematicity, in that participants rejected inferences that resulted from isolated correspondences (despite the fact that such predictions matched well with the base) in favour of those that were supported by a larger causal (i.e. systematic) network. Indeed, the experiment showed that some participants explicitly sought such systematicity in the generation of their inferences.

Experiment 3, in which participants had to rely on their memory representations of the base domain, provided still further (although somewhat weaker) support for systematicity, though the non-availability of the base representation during mapping and inference did result in poorer results than in experiments 1 and 2. However, the results of these experiments does support the hypothesis that systematicity does act as a constraint in the selection process; that the choice of which lower-order relations to map is not determined just by the independent relations themselves, but by the interconnections amongst such relations.

## 2.7.1 Modelling Constraints On Analogical Mapping

Gentner *et al*'s experiments lend credence to the idea that systematicity does act as the main constraint in analogical selection and mapping. As Clement and Gentner argue, "a preference for coherent systems of common information appears to be a psychologically real constraint on analogical mapping". However, whilst the assembled evidence does partially cover some of the criticisms levelled against the structure mapping theory, it does not eliminate the suspicion that the systematicity constraint is just one of a number of constraints acting in analogical transfer. There is a neatness in the representations used in the experiments that rules out the possibility of competing causal networks, and thus the problem of what acts as a constraint in selection on such occasions is avoided. Indeed, if an analog offers up two competing possible modes of transfer with a similar level of systematicity, then the systematicity principle cannot act as a constraint in the selection of one or the other.

One answer to this - mentioned earlier - might be to accept this, but question whether it really is a problem at all. One could simply observe that an analog which leads to such a juncture is a poor choice of an analog (this relates to the earlier criticisms of representations such as Carrol and Mack, 1985). There is considerable psychological evidence that the choice of representation is crucial to analogy (e.g. Holyoak and Koh, 1987). If, when presented with a representation that supports two different, potentially contradictory analogical inferences or mapping networks, the systematicity principle points to two equally valid mappings without preferring either, then this may in fact be a psychologically valid resolution of the initially poor choice of analogy. Indeed, to return to the earlier Holyoak objection to Gentner's structure mapping theory (Holyoak, 1985), it might well be that an expansion of the representations of the solar-system and the atom along the lines he describes will simply lead to a situation in which the two examples are no longer seen to be analogous. Given that the success of any analogy is contingent upon the way in which the putative analogs are represented, it is not a failing of a theory that it cannot provide accurate mappings in a situation where candidate analogs are presented in such a way as to render any analogous similarity between them deeply obscure. Certainly, this would appear to be Gentner's position (Gentner, 1997).[5]

Another approach is to argue that the relations within each description are important relative to some goal (White, 1987), and that this is not captured by the structure mapping theory. White cites the example of Rumelhart and Norman (1981), who used three different analogies to explain the workings of a text editor. The first used the

[5] This point is perhaps best illustrated by considering a *bad* analogy. In 1980, the *New England Journal of Medicine* published a letter from Dr. Stephen E Levick. Dr. Levick's letter noted that over the years, his aluminium cooking pots had become pitted and worn. He explained that the cookware was undoubtedly releasing metal into the food cooked in it. He also noted that the Bantu tribe in South Africa brew beer in iron pots, and absorb so much iron that their cells become filled with it, leading to a variety of ailments. Levick concluded his letter, "The iron pots of the Bantu may have their counterparts in the aluminium pots of industrialised nations, with aluminium-induced dementia [Alzheimer's] as the analogous disease. Large numbers of people in our aluminium-using society may be victims of slow aluminium poisoning from several sources. Corrosible aluminium cookware may be a nontrivial source" (McGee, 1990, p 251). The widespread concern that exists about the use of aluminium cookware is testament to the efficacy of Dr. Levick's analogy. However, had he altered his representation, it may not have worked so well. Tomatoes (extreme takers up of aluminium owing to their acidity) generally contain 1 milligram of aluminium per serving. This increases to 2 milligrams after cooking in aluminium. But this increase pales into insignificance against the 10 milligrams or so of aluminium that anyone who eats raw vegetables will consume per day. Had Levick added these facts to his representation, his analogy with the Bantu may not have proved so memorable; indeed, it might not have even proven to be effective at all.

analogy of a secretary to explain how the editor responded to instructions. However, this analogy led to people thinking that the editor was intelligent, and could distinguish text from commands. A second, using a tape-recorder as an analog for append, caused confusion with regards to text deletion, requiring a third analog, between deletion and card index to convey the idea that when something is deleted, every thing else shuffles up.

White observes that different analogies were required for each explanational sub-goal, and argues that "it appears that the relations within each description will play a part in the analogical process depending on their importance relative to the current goals. If syntactic information alone was used the difference in goals could have no effect on the analogy." (White, 1987, p 22).

However, this doesn't really count as an argument against structure mapping. Indeed, the carry over of the misconceptions that in turn require a further analogy in order to rectify them could be explained as structure mapping in action. In the first instance, the students carry over systematic structures about secretaries irrespective of the goals of their instructors. Thus whilst the goal of removing misconceptions acted as a constraint - for the instructors - in the *selection* of the second analogy, it can be argued that it did not act as a constraint on mapping, and this seems to be backed up by the fact that students transferred a number of plausible relational networks, requiring the third analogy. At each stage the increasing complexity of their view of the text editor - i.e. the relational representation of it - allowed more accurate selections as the effect of the systematicity principle would increase in accordance with the amount of relational information in their text editor representation, and so accordingly, less systematic mappings would be rejected, and a more accurate picture would develop.

In many of the arguments advanced in the literature in support of pragmatics in analogical reasoning, the evidence seems to indicate that pragmatic factors have a clear and demonstrable effect on the selection and representation of analogs. What is less clear is exactly what role (if any) pragmatic factors are supposed to play in mapping. Yet a clear statement of the role that these factors play is essential to the development of a theory of analogical mapping that expresses an effective procedure and is amenable to implementation within a computational cognitive model. At present, the only positive argument in favour of including pragmatics in analogical mapping is that of parsimony: pragmatic factors play a part in analogical reasoning, therefore pragmatic factors play a part in analogical mapping. As I noted earlier, this all boils down to a question of which intuitions a given theorist holds.

Holyoak and Thagard argue that including pragmatic constraints in mapping is the most parsimonious way of reflecting the influence that the evidence shows that these factors have on analogical reasoning. However, from the point of view of developing principled cognitive models of analogy, there may ultimately be a price to pay for this parsimony. In a series of articles Cooper and colleagues (Cooper and Shallice, 1995; Cooper, Fox, Farringdon and Shallice, 1996) have pointed to the need for greater methodological stringency in cognitive modelling, especially in regard to the question of determining exactly how theories relate to models. They argue for a principled distinction between those aspects of a model that are clear embodiments of theory (those aspects that are implementations of an effective procedure) and those aspects of models that are either notation, or are details of the implementation that were required in order to get the model to run. Cooper *et al* call this the A/B distinction: theoretically motivated aspects of a model are described as being *above* the line; implementational details, and specifics of notation, are said to be below it. They argue that cognitive models should make explicit which aspects of a cognitive model are above the line and which are below it if cognitive models are to be amenable to meaningful inspection and analysis, and if the results of running cognitive models are to be treated as meaningful output from theories.

Insofar as pragmatic and semantic constraints are implemented in ACME, it would appear that, in Cooper et al's terms, they are quite clearly B elements. That is, they are below the line of implemented theory. The theoretical motivation behind the inclusion of these constraints has not been stated clearly enough, nor the process that they are supposed to implement sufficiently well-defined, for the semantic and pragmatic constraints to be said to embody theory. Moreover, the semantic links in ACME are quite clearly notation: ultimately, the decision as to whether a strict enforcement of predicate identity is going to be made is in the hands of the person programming the network. In representing the atom <-> solar-system analogy, the programmer can either choose to represent REVOLVES_AROUND as ORBITS in both cases, or else they can use REVOLVES_AROUND in one case, and ORBITS in the other, and then hand code a link between them. In either case, the rationale and the procedure that determines the mapping between the two predicates is clearly external to the ACME model (it is, in fact, in the mind of the programmer).

There is then, a lack of theoretical support for, or a statement of, any theory for semantic and pragmatic constraints in terms of an effective procedure. It follows then, that to date, the only analogical mapping constraints that are both stated as a theory containing an effective procedure, and have supporting evidence for that procedure, are structural constraints. Below, I describe the structure mapping engine, SME, a cognitive model of structure mapping theory.

### 2.7.2 The Structure Mapping Engine

The structure mapping engine (SME; Falkenhainer, Forbus and Gentner, 1989) is a computer implementation of the principles of the structure mapping theory, which simulates the process of interpreting and making predictions from an analogy. SME, uses systematicity extensively: both in evaluating a candidate mapping, and in deriving inferences.

SME first determines all plausible matches between the representations to be mapped by use of a set of *match hypothesis constructor rules*, and the likelihood of any match existing is established by use of *match evidence rules* (these rule sets can be changed to allow SME to be run in various different "modes"). At this stage the program may well have a set of mutually inconsistent matches, with a numerical value representing the evidence for each match.

In the next stage these local matches are collected into *global mappings* - sets of consistent predicate (relation) mappings. SME finds the largest possible sets of matches between predicates, and generates *candidate inferences* for them. The candidate inferences are those object correspondences which must be true for the set match to be plausible. For example, if there is a predicate connected to the base set, but not present in the target system, then that predicate will become a candidate inference in the target.

In the final step, the global matches are considered structurally to see which match has most support in terms of systematic structure. The probability of any global match depends upon:

- the strength of evidence for the individual local matches it comprises;
- the strength of its candidate inference;
- the amount of systematicity (inter-relatedness) in its structure: this is implemented in a trickle-down algorithm in which the evidence for a given

match is increased if there are also matches amongst its parent predicates (Forbus and Gentner, 1989).

SME has been run with apparent success on examples from a number of different domains, (Forbus, Gentner, Markman and Fergusson, 1998; Gentner, Ratterman and Forbus, 1993; Falkenhainer, Forbus and Gentner, 1989). The chief criticisms of these successes have been the restricted domain descriptions upon which SME has operated. In the case of human analogical reasoning, or any possible real world applications to which SME might be put, domain descriptions will be more extensive and complex (though as my earlier remarks indicated, if this were only to lead to SME suggesting a number of - possibly contradictory - mappings, i.e. confusion, then this might not necessarily impugn its claims to psychological plausibility).

From a psychological point of view, the model has three principal features:

- the same processes that are used to form mappings also generate inferences;
- semantics and goals are not required in the formation of coherent matching structures, nor in the generation of inferences
- the model embodies clearly articulated theory only.

As we noted in the discussion above, these features have more psychological plausibility than might first appear.

Just as Gentner's (1989) architecture (described earlier) specifically excluded pragmatic factors from analogical mapping, so Gentner and colleagues specifically and explicitly exclude semantic factors from their model. Thus the actual predicate representations of comparitor objects input to SME are treated as notation. Gentner (Gentner, Ratterman and Forbus, 1993; Gentner Forbus and Law, 1995) assumes that semantic reconciliations take place outside the analogical process: 'SME's constraint of matching identical predicates assumes canonical *conceptual* representations, not lexical strings. Two concepts that are similar but not identical (such as "bestow" and "bequeath") are assumed to be decomposed into a canonical representation language so that their similarity is expressed as a partial identity (here, roughly, "give")' (Gentner, Ratterman and Forbus, 1993 p. 553); conceptual matters are thus explicitly extrinsic to both structure mapping's theory and models (although, it should be noted that Gentner et al provide no clue as to how these decompositions into the canonical representation language take place).

From a cognitive modelling point of view, SME possesses a pleasing simplicity. By focusing upon only one factor in mapping, namely the one that evidence shows is the factor that is most important to - if not the only factor in - analogical mapping, it offers the possibility of gaining an insight into the influence of that factor in the mapping process. Whereas Holyoak and Thagard try to model every possible constraint leading to mapping, creating a muddied picture of the role of individual factors, SME can also enable one to gauge the role of factors other than structure by measuring any performance gap between the purely structural SME implementation of analogical mapping and other models which embody other theoretical constraints.

## 2.8 Comparing The Two Theories And Models Of Mapping

It would seem, from the foregoing, that many of the theoretical criticisms of Gentner's structure mapping theory result from the problematical nature of analogical reasoning, rather than from any deficiencies in Gentner's account of it. Moreover, there are real problems in giving an adequate account of the factors, included within Holyoak's constraint satisfaction theory, which have been proposed in order to provide an account of analogical mapping that incorporates a "richer" notion of semantics and pragmatics.

As the discussion of the models of the various theories showed, there is a great deal of convergence between SME and ACME, to the extent that it can be argued that ACME is in actuality more of a model of structure mapping theory than it is of constraint satisfaction theory (at least in so far as the latter is proposed by Holyoak and Thagard), since influence of constraints other than structural considerations is marginal (and arbitrarily controlled within the model).

None of this counts as evidence against or in favour of either theory as an account of analogy. However, since both camps claim experimental support for their models, and since both models are ultimately structure mappers, it would appear that, to date, what evidential support models of analogy do provide is in favour of Gentner's structure mapping theory. The role and influence (if there is any) of specific mapping constraints other than those which rely upon structural information remains largely unexplored in any theoretically precise way.

# 2.9 Accessing Analogies

## 2.9.1 Theories Of Retrieval

The question of accessing analogies is closely bound up with judgements of similarity. Analogy is merely one of a number of ways by which two things might be adjudged similar. Accordingly, in general, accessing models for analogy are based upon the more general principle of similarity based retrieval.

Gentner, Ratterman and Forbus (1993) distinguish differing categories of similarity match:

- *Literal similarity* matches include both common relational structure and common object descriptions;
- *Surface matches* are based upon common object descriptions, with some shared first order relations;
- *Analogy*, as described earlier, a match based upon a common system of internal relations.

Thus this account, in keeping with earlier work, defines similarity in terms of degrees of correspondences between structured representations. The new feature, however, is that Gentner and Forbus argue that similarity based access from long-term memory relies more on surface similarities and less upon structural commonalities than mapping. Gick and Holyoak (1980) observe that people often fail to access potentially useful analogs, whilst Ross (1984; 1987) showed that whilst people engaged in problem solving are often reminded of prior problems, these remindings are usually based on surface rather than structural similarities between solution principles.

Holyoak and Thagard accept Gentner's argument for the influence of semantic similarity on the retrieval process. However, whilst they accept that semantic similarity acts as the dominant constraint in retrieval, they argue that it does not act alone.

Thus Holyoak and Thagard carry their template of constraint satisfaction over to their model of retrieval, i.e. they see three basic kinds of constraint acting as pressures on the selection process: semantic similarity, isomorphism (informally treated, so that it refers to structural similarity) and pragmatic considerations.

**Base Story**

Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so grateful that he pledged never to shoot at a hawk again. He went off and shot a deer instead.

**Literal similarity**

Once there was an eagle named Zardia who nested on a rocky cliff. One day she saw a sportsman coming with a crossbow and some bolts that had no feathers. The sportsman attacked but the bolts missed. Zardia realised that the sportsman wanted her tailfeathers so she flew down and donated a few of her tailfeathers to the sportsman. The sportsman was pleased. He promised never to attack eagles again.

**True Analogy**

Once there was a small country called Zardia that learned to make the world's smartest computer.

One day Zardia was attacked by it's warlike neighbour, Gagrach. But the missiles were badly aimed and the attack failed. The Zardian government realised that Gagrach wanted Zardian computers so it offered to sell some of it's computers to the country. The government of Gagrach was very pleased. It promised never to attack Zardia again.

**Figure 2.5 (part 1):** Example stories from Ratterman and Gentner (1987).

**Mere Appearance - With First Order Relations**

Once there was an eagle named Zardia who donated a few of her tailfeathers to a sportsman so he would promise never to attack eagles.

One day Zardia was resting high on a rocky cliff when she saw the sportsman coming with a crossbow. Zardia flew down to meet the man, but he attacked and felled her with a single bolt. As she fluttered to the ground Zardia realised that the bolt had her own tailfeathers on it.

**False Analogy**

Once there was a small country called Zardia that learned to make the world's smartest computer. Zardia sold one of its supercomputers to its neighbour, Gagrach, so Gagrach promised never to attack Zardia.

But one day Zardia was overwhelmed by a surprise attack from Gagrach. As it capitulated the crippled government of Zardia realised that the attacker's missiles had been guided by Zardian supercomputers.

**Mere Appearance (Attributes only) - Ratterman and Gentner, Experiment 2**

There once was a sportsman who loved to hunt. He liked to have the animals he caught stuffed and mounted. His pride and joy was an eagle he had killed with just a crossbow and a bolt. He had been hiding in the top of an elm tree when he shot her.

**Figure 2.5 (part 2):** Example stories from Ratterman and Gentner (1987).

Gentner and Landers (1985) and Ratterman and Gentner (1987) (see also Gentner, Ratterman and Forbus, 1993) examined the role of structural and semantic factors in access. Gentner and Landers' experiment had two purposes:

1. to test the prediction that shared systematic structure determines the subjective soundness of a match;
2. to see whether the accessibility of analogies (and other similarity matches) mirrored their inferential soundness.

| | | Higher Order Relations | |
|---|---|---|---|
| | | Shared | Not Shared |
| Object Attributes | Shared | Literal Similarity | Mere Appearance |
| | Not Shared | True Analogy | False Analogy |

**Table 2.1:** Breakdown of story attributes.

The study was designed to recreate the conditions of natural long-term memory access: participants were given approximately 30 stories to read and remember; one week later they read a new set of stories, reporting any cases where a new story reminded them of any of the originals. The stories were designed to embody three basic kinds of similarity match:

i. mere appearance (MA) - the base and targets shared object descriptions, for instance a hawk in the base related to an eagle in the target and first order relations (i.e. ATTACKED(x,y) for SHOT_AT(x,y);

ii. true analogy (TA) - the base and target shared first order relations and higher order relations (e.g. relations between relations: CAUSE(S(x,y),R(y,z)) in the base was CAUSE(S'(x',y'),R'(y',z')) in the target);

iii. false analogy (FA) - where only first order relations matched.
(See also figure 2.5 and table 2.1).

Gentner and Landers found that the proportion of mere appearance matches was far greater than the true analogy matches, which in turn outnumbered the recalled false analogy matches. This result suggested that surface commonalities significantly influence memory access. However, when the same participants were then asked to rate the inferential soundness of their matches, the true analogy matches were rated to be significantly more sound than the MA and FA matches.

The Gentner and Landers study threw up an interesting insight - it showed that the kinds of matches that participants consider most sound were not the kinds of match that most promote access - but it raised a number of questions for which it offered no solutions. As well as suggesting that surface commonalities had a significant role to play in memory access, it also seemed to show that higher order relations played some role, since false analogy matches were retrieved much less than true analogies. The results obtained did not explain how, or even whether, surface attributes and higher order relations combine to promote access, nor did they isolate the aspects of surface similarity which most significantly influence access.

Ratterman and Gentner set up a number of experiments to address some of these questions. The first was intended to investigate the respective roles of surface commonalities and higher order relations in access. In order to do this, they replicated the Gentner and Landers study, adding another match type:

iv. literal similarity (LS) - the base and target have commonalities at all levels (see also figure 2.5 and table 2.1).

From the results of the Gentner and Landers study, Ratterman and Gentner expected that LS matches would be recalled frequently in the reminding task.

One problem with the Gentner and Landers study is that it did not eliminate the possibility that retrieval was in fact simply dependent upon the overall similarities between the matches (i.e., the MA matches were just more similar to one another than the TA and FA matches). In order to clarify this, Ratterman and Gentner added a similarity rating task in order to test whether retrieval could be predicted simply by similarity ratings. Thus Ratterman and Gentner were testing 3 parameters:

1. accessibility (recall)
2. inferential soundness
3. the degree of similarity between base and target (see table 2.1).

By repeating the Gentner and Landers study in this way, Ratterman and Gentner discovered:

1. accessibility: literal similarity and mere appearance led to significantly more remindings than true analogy and false analogy, supporting the results of the Gentner and Landers tests.
2. soundness rating: again the results of the Gentner and Landers tests were verified, with the TA and LS matches being considered significantly more sound.
3. similarity rating: participants rated the LS matches significantly higher than the TA matches, and both significantly higher than MA and FA, which were given equally low ratings. Moreover, the pattern of similarity ratings mirrored those for soundness ratings.

Since the results of the recall task on the one hand, and the similarity and soundness ratings on the other, varied markedly, Ratterman and Gentner concluded that different aspects of similarity govern the different processes, with surface matches playing the major part in recall, and structure being most significant in judging soundness and similarity.

A second experiment was designed to further analyse the role of surface commonalities, examining which aspects of the mere appearance matches lead to their accessibility, and whether object descriptions could promote access on their own. A second set of mere appearance matches were created (see figure 2.5), in which all the first order commonalities from the first set were removed. The study was then run again, with the new mere appearance (attribute only - MAAO ) matches included along with the first MA set (renamed MAF - mere appearance (first order)), LS and TA matches.

In the second experiment recall test, the LS and MAF matches were recalled significantly more than the MAAO matches, which in turn were recalled significantly more than the TA matches (recall of both the latter sets was poor). As with experiment 1, this ordering was in contrast to participants' opinions as to soundness and similarity.

Ratterman and Gentner concluded from these experiments that whilst similarity based recall is not *only* influenced by common object descriptions (since the biggest gain in recall comes from object descriptions and first order relations) there is a clear dissociation between the kinds of similarity that are inferentially reliable and the kinds of similarity that enable memory access.

Accordingly, the Gentner et al similarity based transfer process is decomposed into two sub-processes that are qualitatively different

- *Accessing* a similar (*base*) situation in long-term memory, based primarily upon surface similarity
- Creating a *mapping* from base to target using structural commonalities.

Holyoak and Thagard's conjecture is that the role of isomorphism in retrieval might be more important than Gentner supposes. As Gentner notes (Gentner and Forbus, 1991), her findings were not that structural similarities do not contribute to retrieval, but rather that such similarities have a much greater effect on mapping once two analogs are present than they do on similarity based retrieval. (Ratterman and Gentner (1987) had observed that structural similarities led to a non-significant increase in retrieval in two studies, and small but significant increase in a third.)

The key point of note here is that according to Gentner, analogical access is probably based upon qualitatively distinct processes from analogical inferencing (in which she includes mapping). Holyoak and Thagard's conjecture is based upon their disagreement with this notion, and upon their corresponding re-emphasis of the accepted (albeit slight) role played by isomorphism in retrieval. They do not present any new evidence that isomorphism plays a greater role than that noted by Gentner. Indeed, the results of Holyoak and Thagard's own experiments with retrieval implied that:

> "semantic similarity ... dominates reminding for novice candidates"
> (Wharton, Holyoak, Downing, Lange, Wickems and Melz, 1994, p 98)

Their argument for the importance of pragmatic considerations (goals) consists of the following: the purpose of analogy is to help accomplish the goals of the problem; clearly a retrieval system attuned to increase the retrieval of analogs relevant to goal accomplishment would contribute more to effective problem solving than a retrieval system that lacked sensitivity to goals. Again, however, no direct psychological evidence is offered to support this. Indeed, as Thagard et al (1990) note, a problem in this area is that it is difficult to distinguish effects of goals that reflect particular constraints from the effects that can be attributed to consequences of other constraints such as structural similarity.

Indeed, it could be argued that the view of problem solving espoused by Holyoak and Thagard is particularly vulnerable in this area. Firstly, it seems impossible to give an account of the role of pragmatic factors in analogy that could not be explained equally in terms of mappings between structures in the problem definition. Moreover, the specification of goals is a key component of problem solving; in some accounts, defining the problem is what problem solving is. In many cases analogy is used in order to assist in the definition of goals (Holyoak and Koh, 1987). In such cases a structural explanation will be less open to criticisms of circularity than one in which a goal (to find a problem definition?) is used to select an analog which will help accomplish the goals of the problem (provide a problem definition?).

## *2.9.2 Implementations Of Theories Of Retrieval*

The two competing theories of analogical access described above have been implemented as computer programs:

## *2.9.3 ARCS - Analogical Retrieval By Constraint Satisfaction (Holyoak And Thagard)*

Holyoak and Thagard have implemented their ideas in an experimental system, ARCS (Thagard, Holyoak, Nelson and Gochfeld, 1990). ARCS uses a localised connectionist network in order to apply semantic, isomorphic (structural) and pragmatic constraints to selecting potential analogs from memory. The initial stage of the algorithm (figure 2.6) uses semantic constraints to form a suitable subset of memory over which the matching network is built. The concept of semantic similarity is based upon a mixture of a thesaurus and WordNet (Miller, Fellbaum, Kegl and Miller, 1988), a database of words and lexical concepts, which are used for determining semantic similarity between the stored words - in this case the vocabulary of the predicate representations.[6]

The procedures of ARCS are carried out by use of a constraint satisfaction network similar to ACME (described in section 2.6.3, above). The settling of the network provides an ordered set of retrieval hypotheses, based upon the activation levels at their nodes.

---

Given a pool of memory items $I_1..I_n$, and a probe P,

1. For each item $I_i$, include it in a matching network if there are any similar predicates in $I_i$ which are semantically similar to predicates in P. The matching network implements semantic and structural constraints.
2. Create inhibitory links between units representing competing retrieval hypotheses to ensure competitive retrieval.
3. Install pragmatic constraints by creating excitatory links between a special pragmatic node and every predicate marked by the user as important.[7]
4. Run the network until it settles.

---

**Figure 2.6:** The ARCS Algorithm (summary taken from Law, Forbus and Gentner, 1994)

## *2.9.4 MAC/FAC*

MAC/FAC (for "many are called, few are chosen"; Gentner and Forbus, 1991) uses a two-stage retrieval process based upon Gentner's retrieval theory. It comprises MAC, a crude, computationally cheap matching process used to select a limited number of candidates for more expensive matching using FAC (which is SME, described in

---

[6] Whilst this hardly amounts to a scientific account of the way semantic links are managed, it does have the beneficial effect of replacing semantic links dependent upon the idiosyncratic, *ad hoc* assumptions of the network programmer with semantic links based upon a standardised set of *ad hoc* assumptions.

[7] Thagard *et al* accept that the psychological evidence for this node is limited - 'part of the problem is that it is difficult to distinguish effects of goals that reflect a special pragmatic constraint from effects that can be interpreted as consequences of other general constraints involving semantic similarity and structural consistency... as in ACME... we treat pragmatic factors as an additional pressure to semantic similarity and structural consistency' (Thagard *et al*, 1990, p 267).

section 2.7.2), in order to apply structural constraints to select the best match(es) (see fig 2.7).

Since there is little agreement as to the structure of long-term memory, MAC/FAC assumes only that there is a stage at which a limited number of memory items[8] are selected from a pool according to their similarity to a probe (target). This memory item (or items) is a potential base analog. Both the MAC and FAC stages consist of a *matcher*, which is applied to every input selection, and a *selector*, which uses the evaluation of the matcher to select which comparisons are produced as the output to that stage.

---

Given a database M of memory items $I_1..I_n$, and a probe $P$,

1. [MAC Stage] In parallel, for each item I in M compute the dot product of the content vectors for $I$ and $P$. Return as output the maximum and every item whose score is within *p1%* of it.
2. [FAC Stage} In parallel, for each item I in the MAC output, run SME with $I$ as the base and $P$ as the target. The FAC score for each pair is the structural evaluation score of the highest ranked mapping. The top-scoring match, plus any others within *p2%* of it, are output.

(Typically, $p1 = p2 = 10\%$)

---

**Figure 2.7** The MAC/FAC Algorithm (Law, Forbus and Gentner, 1994; Forbus, Gentner and Law, 1995)

The MAC stage matcher is used to estimate how well FAC will evaluate comparisons, in order to filter down candidates into a number suitable for the more extensive (and computationally expensive) processing in FAC.

MAC computes the *numerosity* of the comparison, that is, it estimates the number of match hypotheses that FAC would generate in comparing the probe to a given memory item. If few local matches are hypothesised, then the best global interpretation can be assumed to be small. However, numerosity is not a perfect measure, since having a large number of local matches is no guarantee of a large global match (for instance match hypotheses may be ungrounded because of a lack of correspondence in their

[8] Some kind of conceptual 'instance'.

arguments). The MAC process calculates numerosity by use of *content vectors*, which are n-tuples, each component number of which corresponds to either a predicate, a function or a connective in a structure description[9]. MAC matches by introducing a probe, whose content vector is computed. A score is generated for each item in the memory pool by taking the dot product of its content vector with the probe's content vector. These scores are sent to the MAC selector, which returns as output the best match, and everything within *n%* of it.

The FAC stage computes literal similarity according to structure mapping, implemented using SME, currently with appropriate rule sets to map literal similarity rather than analogy. Literal similarity is used rather than analogy in order to achieve the highly observed frequency of surface remindings, most of which would be rejected if FAC were strictly an analogy matcher.

Since MAC is only sensitive to predicate overlap, and FAC is structure sensitive, FAC will reject much of MAC's output; the filtering provided by MAC acts to cut down the number of matches FAC is required to do.

## 2.10 Discussion

### 2.10.1 Comparing The Two Theories And Models Of Processing

The divergence between the theoretical models of analogical processing appears to be wider than the gap between implementations. Whilst the structure mapping theoretical model embraces two distinct processes, one for analogical accessing, and one for mapping and inference, the multi-constraint theory position is that a single account can

[9]• P = the total set of functors (predicates, functions, connectives) used in the descriptions that constitute pool items and probes.

• a *content vector* = an n-tuple of numbers, each component corresponding to a particular member of P.

• Given a description D, the value of each component of its content vector indicates how many times the particular corresponding member of P occurs in D.

• Members of P not present in D are assigned the value 0.

Thus, for example, if there were four instances of IMPLIES in a story, then the value for IMPLIES in its content vector would be 4.

be given. When these models are then implemented, this gap narrows, since it is debatable whether ARCS amounts to two processes in practice.

As mentioned in discussion earlier, this divergence between theory and practice should cause some alarm to the multi-constraint theory advocate, since if the theory is right, it cannot be correct to claim that the output from the implementation fully supports it, at least in a competition with the structure mapping theory. In practice the success of ACME in simulations, given its structure dominated mappings, offers more support for the structure mapping theory than it does for multi-constraint theory. This leaves the theoretical implications of multi-constraint theory unproven. Constraint satisfaction might indeed be a better model of the analogical process, but whilst ACME practices what is in effect structure mapping, and the accessing element in ARCS uses mainly semantic considerations, then positive results from experiments will add to the psychological plausibility of Gentner's two stage model rather than multi-constraint theory.

On the other hand, in defence of multi-constraint theory, constraint satisfaction isn't a *single* process; it embraces a number of processes, including those described within Gentner's model. As Ratterman and Gentner (1987) acknowledge, the evidence implies that accessing does involve structural (first order) relations as well as the simple surface commonalities utilised by MAC/FAC. Gentner's two process model fails to capture this constraint, whilst Holyoak and Thagard's model does. Yet according to both theories there is very good evidence for supposing that first order relations do play a role in accessing.

This divergence can be reconciled to some extent if we see the two theories as offering different granularities of description. According to this view, Gentner's theory gives a coarse grained description of analogy, focusing upon the main process in a particular analogical sub-process, whereas Holyoak and Thagard's is a finer grained model, aimed at capturing all of the constraints that influence analogical reasoning, albeit, perhaps, at the expense of the conceptual clarity provided by an overview.

On this view, the Holyoak and Thagard, and the Gentner models are best seen as different but complementary aspects of a notional larger model, rather than competing and contradictory versions. The evidence suggests not that there are two totally distinct analogical processes, but rather that two different constraints play the dominant role depending upon the task in hand. The Holyoak and Thagard model is based upon the former notion, whilst the Gentner model captures the latter (see figure 2.8).

## Gentner's Structure Mapping Theory

| | Theory | Model | Evidence |
|---|---|---|---|
| **Mapping** | | | |
| Structural | √ | √ | √ |
| Pragmatic | x | x | ?? |
| Semantic | x | x | ?? |
| **Retrieval** | | | |
| Semantic | √ | √ | √ |
| Pragmatic | x | x | x |
| Structural | √ | x | √ |

## Holyoak and Thagard 's Multi-Constraint Theory

| | Theory | Model | Evidence |
|---|---|---|---|
| **Mapping** | | | |
| Structural | √ | √ | √ |
| Pragmatic | √ | ? | ?? |
| Semantic | √ | ? | ?? |
| **Retrieval** | | | |
| Semantic | √ | √? | √ |
| Pragmatic | √ | ? | x |
| Structural | √ | √ | √ |

**Figure 2.8:** The main points in each theory summarised in tabular form.

## 2.11 Implications For The Cognitive Modelling Of Analogy

### 2.11.1 Retrieval

Neither model of retrieval matches perfectly with the evidence. MAC/FAC doesn't capture all of the constraints which the evidence suggests influence analogical retrieval, as it does not model the role that first order relations were found to play in recall by Gentner, Ratterman and Forbus.

ARCS can model the structural constraints on retrieval which receive support from the psychological evidence, but in addition, it also models constraints which receive minimal (i.e. speculative) or non-existent support as well.

At present, it would appear that if MAC/FAC could be extended in order to account for the role of first order relations in retrieval, then it could broadly model all of those constraints which the evidence indicates are important influences on analogy (experimentation with such a model might be useful in determining how close the available evidence comes with respect to painting a complete picture of analogy - for instance, the relative performance of a model like this might ultimately be useful in determining the necessity and plausibility of the extra constraints proposed in ARCS).

### 2.11.2 Mapping

#### 2.11.2.1 Structural Factors

As far as mapping goes, structural constraints receive support from both theories, and from the great deal of empirical work that has been done exploring analogy. Both models employ structural constraints to determine analogical mapping, although these constraints differ in some details. These structural constraints are neatly summarised by Keane, Ledgeway and Duff (1994) as follows :

- *Matches only between entities of the same type*: attributes are matched with attributes, objects with objects and two-placed predicates with two-placed predicates. So in *foo(A,B)*, the predicate *foo* would never be matched with the object *C*.

- *Structural consistency*: if the propositions *foo(A,B)* and *foo(C,D)*, are matched, then their arguments are matched correspondingly (i.e. $A$ with $C$, and $B$ with $D$).
- *Favouring systematic matches*: when faced with alternative sets of matches, the mapping with the highest-order of connectivity should be preferred.

Holyoak and Thagard (1995) argue that the similarities and differences between the structural constraints in their theory and those posited in structure mapping theory can be best illustrated by comparing Gentner's theory with the three main constraints posited by multi-constraint theory. They claim that multi-constraint theory captures Gentner's insight regarding the importance of systematic structure (in ACME, interconnected systems will have more mutually supporting links than an isolated relation), but in a more flexible manner. As implemented in SME, Gentner's theory rigidly enforces one-to-one mappings and structural consistency — potential mappings which violate these constraints are not made. In contrast, ACME, whilst preferring one-to-one mappings (by using inhibitory links to discourage many-to-one mappings) nevertheless will allow violations (in ACME, some structural constraints are seen as 'soft' constraints).

Moreover, because both theories clearly express their structural constraints as effective procedures, it is possible to carry out empirical studies in order to ascertain which particular structural constraints are to be preferred. Spellman and Holyoak (1992), found that a minority of participants in their study did make many-to-one mappings (as allowed for in multi-constraint theory but not in structure mapping theory). In making an analogy between the start of the Gulf War and the start of World War II, 9% of participants mapped Kuwait onto two or more of Austria, Poland and Czechoslovakia, whilst a few made mappings between Kuwait and things like "countries Hitler invaded", re-coding the multiple items to preserve one-to-one mappings.

It was not possible, given the way Spellman and Holyoak's study was conducted, to determine whether participants were in fact making many-to-one *mappings*, or whether they were making a series of isomorphically consistent one-to-one mappings, and reporting an overall analogical mapping based on an aggregation of these individual mappings; or indeed, whether participants were grouping objects (as described above) prior to mapping, computing a one-to-one mapping, and then reverting to a representation with multiple items in reporting their analogical mapping (Spellman and Holyoak, 1997).

In recent work, Markman (1997) has provided evidence to show that whilst analogical inferences are heavily constrained by structural factors - as both theories predict - in studies where materials allow for potential many-to-one mappings in generating inferences, participants appear only to make one-to-one mappings between individual analogs, even when they go on to report overall analogical mappings based on aggregations of individual mappings.

Whilst the question of whether one-to-one mappings should be considered as a hard constraint in analogical mapping has not been settled for certain - there is still a lot of space for further empirical study - it should be stressed that it is the fact that both theories express the structural constraints they propose in terms of detailed effective procedures that makes this kind of subsequent empirical validation possible.

#### 2.11.2.2 IAM (The Incremental Analogy Mapping Engine)

Another theory of analogical mapping that shares many features with SME and ACME (and also has a base in empirical studies of human analogising) is modelled in IAM (Keane and Brayshaw, 1988; Keane, Ledgeway and Duff, 1994). In addition to possessing many interesting features in its own right, in respect of the three constraints discussed here, IAM represents a kind of 'half-way house' between SME and ACME, since it enforces a very similar structural constraint to SME, whilst also allowing the influence of semantic and pragmatic constraints in the manner of ACME.

IAM's structural constraints are implemented according to the summary above, however, the manner of their implementation differs widely from SME (Keane and Brayshaw, 1988; Keane, Ledgeway and Duff, 1994). In IAM, a mapping is formed from a sub-set of the predicates / elements in a base domain, rather than from a computation amongst all such elements (as carried out by SME).

IAM selects the group of predicates that have the greatest higher-order connectivity between its elements (the *seed-group*), and, having done this, it chooses an element from the seed-group and finds the best match between this element and all the elements in the target (the *seed-match*). This seed-match is then used to 'grow' the mapping of all of the elements in the seed-group: all of the legal matches between the remaining elements in the seed-group and the target are found, and a set of unique set of one-to-one matches is formed. Like ACME, IAM uses serial constraint satisfaction to find these matches, applying structural, pragmatic and semantic-similarity constraints. Again, as with ACME, pragmatic and semantic-similarity constraints are realised by manipulating weightings in the network. However, rooting the search in one element - IAM favours relational elements taking multiple objects as seed-elements - and resolving ambiguities produced in mapping by reference to that element, enables IAM to produce one-to-one matches of the kind produced by SME, and avoid the many-to-one matches that can sometimes be generated by ACME, in a manner that is particularly efficient computationally.

Depending on task demands, and the success of the matches grown from an initial seed (Keane et al report that assessing whether more than half of the seed-group predicates have been matched successfully proves to be sufficient in gauging this), IAM may either backtrack and try other seed-matches from the seed-group, or else, if none of the mappings produced is optimal, go on to map other groups of predicates in the base domain (although Keane *et al* 1994 note that in practice this latter step tends not to be necessary).

Keane, Ledgeway and Duff (1994) report that IAM has successfully emulated the analogical performance reported by SME and ACME in simulations (Falkenhainer et al, 1989; Holyoak and Thagard, 1989). Moreover, Keane et al argue, with evidence, that the efficient nature of the algorithm used to implement IAM is to be preferred on behavioural grounds to those employed in SME and ACME.

#### 2.11.2.3 Pragmatic Factors

As noted earlier, whilst there is considerable evidential support for the role that pragmatic factors such as goals and plans play in the overall analogical reasoning process, there is no evidence to decide the question of which of the in-mapping or pre- and post- mapping accounts is correct (and where evidence has been sought to clarify this, it has failed to do so, e.g. Spellman and Holyoak, 1992; 1996; Keane, 1988; 1996). This problem is best illustrated by considering the way that pragmatic constraints are implemented in models that incorporate them, such as ACME or IAM: in both of these models, predicates or nodes that are considered pragmatically useful are highlighted in the representation of the analogy, and this highlighting is used to focus mapping. However, no theoretical claims are made for the particular way that representations are highlighted (the markers are simply implementational details), and no theoretical claims about the way pragmatic markers are processed. Thus it seems fair to say that from a theoretical standpoint, there is nothing to distinguish a highlighted structure in ACME or IAM from a decision to highlight a structure in SME by embedding it within a higher order structure that described its pragmatic usefulness. In either case the effect would be the same, and in either case, from the behaviour of the models, it seems more accurate to see pragmatics as a constraint acting upon *representation building* rather than *mapping* (or at least, there seems to be no theoretical or evidential evidence to prefer an account of pragmatics acting directly upon the mapping process).

Since there is no in-mapping account that can be described in terms of an effective procedure, and since structural factors in mapping receive such massive theoretical and empirical support, at least one advantage of a or pre- and post- mapping account from a cognitive modelling perspective is that it enables a model of analogical mapping to be specified with A/B factors clearly and explicitly noted, whereas this will not be the case for a model which tries to implement pragmatics without any clear, operationalised idea of how pragmatic factors influence analogical mapping. Again, however, this is clearly an area where further study is needed.

#### 2.11.2.4 Semantic Factors

Perhaps the most interesting area of debate in analogy - from a wider cognitive perspective - is the role semantic factors play in analogy. In the introduction to this chapter, I examined the way analogy is typically characterised in the cognitive science literature. Typically, analogy is defined as some kind of mapping that permits two concepts to be seen as similar on the basis of some relational similarity that holds between them.

Theories differ in that structure mapping theory assumes that the process that determines conceptual reconciliation between semantically similar predicates is external to the analogical mapping process, whereas multi-constraint theory posits some kind of interaction between the two. However, neither theory subscribes to a particular theory of semantics, or conceptualisation.

Clearly, if one could specify a categorisation model that could act to determine conceptual reconciliations during the on-line processing of analogies, then this could act as a module that could be added to either model in order to determine empirically which is more accurate. On the other hand, given the way that analogy is characterised, *failure* to specify exactly what counts as a particular concept - so that literal similarity comparisons can be distinguished from analogical ones - might have deep theoretical consequences for theories of both analogy and categorisation.

## 2.12 Whither Analogy?

Ordinarily we accept distinctions between category membership and analogy according to the tacitly realist terms used in the definitions of analogy described earlier. In categorical judgements, relating a new representation of an object to some kind of stored category representation, objects are felt to be similar to one another in a way in which those objects in judgements of analogical association are not. If two objects are considered to be members of a category, the classification is real; if they are considered to be analogous, it is not. Consider, for example an analogy between a theory and a building (Lakoff and Johnson 1980): we might talk of "the *foundations* of a theory"; "we might wish to *buttress* a theory with more facts"; "theories that we *construct* can also *collapse*". From an everyday, psychologically realist viewpoint, an igloo and a castle and a skyscraper really are similar in a way that buildings and theories are not.

The research into analogy described above has accepted this tacit realism, as indeed has much research into metaphor. Holyoak and Thagard (1995) describe a world in which "we think we see things as they really are", and analogy is used in order to recycle our existing knowledge of the real world to formulate new bits of 'real' knowledge. (Similarly, in the case of metaphor, Ortony (1979) makes a distinction between literal and non-literal similarities: 'encyclopaedias are like dictionaries' is true in a literal (real) way, whereas 'encyclopaedias are like goldmines' is only true in a metaphorical (non-real) way. Whether the notion of literal similarity might be problematic or not is barely examined, since the real problem to be addressed is metaphor.)

Analogies, in the research described above, are defined as being distinct from categories, the nature of which are left unexamined. Tacitly, categories are presumed to be real. But without an account of categorisation, distinctions between analogy and metaphor reliant upon a contrast with categorisation fail to do any distinguishing at all. Analogy is consistently defined in contrast to categorisation (Clement and Gentner, 1991; Holyoak and Thagard, 1995). In order to make a contrast definition one needs an account of at least one of the contrasting elements. Without an account of categorisation, a definition such as:

> "In an analogy, a familiar domain is used to understand a novel domain in order to highlight important similarities between the domains, or to predict new features of the novel domain." (Clement and Gentner, 1991, p89)

might be more accurately reformulated along the lines of: 'in analogy, a stored representation is used in order to highlight important similarities between it and a new representation of an object or concept, or to predict new features in the new representation of an object or concept', none of which would be out of place in a definition of categorisation.

As I described in section 2.2, the distinction between categorisation and analogy is difficult to draw. In the following two chapters, I shall argue that an account of categorisation sufficient to make the kind of distinctions put forward in the cognitive science literature cannot be given. Following from that, in the rest of this thesis, I shall go on to explore the hypothesis that at cognitive levels of description there may be no clear distinction to be made between analogy and categorisation at all.

# Chapter 3

# Categorisation I
# Wittgenstein's Theoretical Challenge

> *66. Consider for example the proceedings we call "games". I mean board-games, card-games, ball-games, Olympic-games, and so on. What is common to them all? - Don't say: "There must be something common, or they would not be called 'games'" - but look and see whether there is anything common to all. - For, if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat, don't think, but look!*
> Wittgenstein (1953, §66, p31).

## 3.1 Wittgenstein, Concepts And Categories

Although we perceive the world through our senses, we do not inhabit a world of sense impressions. Rather, as Indurkhya (1997) very vividly describes, we live in a world of concepts:

> "When I look around, I do not 'see' the patterns of activity of the retinal cells in my eyes; I see trees, birds and rivers... When we act on our environment, we act upon this world of concepts. We plant a tree, bathe in a river, move a log across a field, etc." (Indurkhya, 1997, p. 8)

Most, if not all, cognitive activities involve this process of converting the mass of data we receive from our senses into 'meaningful' concepts. This process works - in broad terms - by grouping, or clustering sense impressions into larger wholes (or representations). As Indurkhya puts it, it works "by reducing the myriad of sense impressions into a handful of meaningful concepts and categories" (Indurkhya, 1997, p. 8). Although any given dog will form a different image on the retina, or even the same dog at different times, still we are capable of grouping all these sensations together under a single label - 'dog'.

We can also *impose* concepts on sense data. As gestalt psychology has shown, when sensory stimuli comprise two flashes of a dot in close proximity, what we 'see' is a smooth and continuous motion of the dot from the first position to the second position (Kolers, 1972; Kolers and Green. 1984). Similarly, our sensory data can be deconstructed, for instance, in the way that continuous streams of sound are separated into individual words when a listener parses sentences.

Concepts are a vital aspect of cognition. As with analogy, cognitive science has a strong interest in categorisation. Accounting for how the 'stuff of experience' is represented, manipulated and combined in the mind is a central concern of many researchers in the field. Moreover, as can be quickly gleaned from even the most casual perusal of the relevant cognitive science literature, Wittgenstein's analysis of concepts and categories in the *Philosophical Investigations*[1] (1953) has had a great influence on the approaches taken in this area.

Before reviewing the current state of the art with regards to the cognitive modelling of categorisation in the chapter following, I will in this chapter lay out a theoretical framework within which the question of modelling concepts and categories can be considered. At the beginning of my review of cognitive models of analogy, I discussed the lack of any clear working characterisation of analogy in the cognitive science literature, and sketched out a rough map of the area covered by the concept of analogy as it is used in the cognitive science literature. Here, I shall attempt to develop a similar framework - though, for reasons that will become clearer as I develop this position, not a definition - for exploring categorisation. This framework is based on an analysis of concepts and naming presented by Wittgenstein in the *Philosophical Investigations* (1953).

In spite of the powerful influence of Wittgenstein's analysis, no clear statement of his position appears in the cognitive science literature. It is worth noting that, Backer and Hacker's (1980) fine, albeit slightly impenetrable exposition excepted, the relevant sections of the *Philosophical Investigations* have been surprisingly little discussed in the philosophical literature. Indeed, a recent guide to the *Philosophical Investigations*

[1] For the sake of convenience, I shall often abbreviate *Philosophical Investigations* to PI in the discussion that follows.

aimed at academic philosophers (M. McGinn, 1997) completely ignores these sections altogether.

It is worth stating at the outset of my attempt to present a detailed exposition of Wittgenstein's arguments relating to concepts and categories, that 'concepts' and 'categories', as they are studied in cognitive science were not the focus of Wittgenstein's investigation; rather, Wittgenstein's concerns ran somewhat parallel to those of cognitive science's in this regard. In this chapter, I shall endeavour to show that Wittgenstein would have rejected the conception of concepts that he is held to have put forward, as indeed he would have rejected the very conception of 'concepts' as they are generally treated in cognitive science. Therefore some elaboration will be required in order to bring these two views of concepts together so that they might be compared and contrasted.

Wittgenstein is often presented as an opaque, difficult to interpret, and rather obscure philosopher, which sometimes leads to the *Philosophical Investigations* being seen as - and treated as - a philosophical pick 'n' mix; a series of gnomic quotables to be plundered in support of a thesis. As a philosopher, Wittgenstein can be seen as a mirror, in whose writing a reader can see what she wishes to see. I shall endeavour to show, by examining in detail the arguments that Wittgenstein presents, that PI sections §66 to §82 lay out a clear, if intricately connected, series of arguments, which, with a certain degree of elucidation, can be seen to systematically detail Wittgenstein's theoretical treatment of categories and categorisation in what is - for Wittgenstein - a fairly straightforward manner. What emerges from the close reading of Wittgenstein's text that I present is at considerable variance with the account of Wittgenstein's position that has been generally accepted in the literature.

There are a number of reasons for this, perhaps the most important being that whilst Wittgenstein is often cited as a founding influence in cognitive approaches to concepts and categorisation, his concerns - and more importantly his methods - were markedly different to those of researchers in the modern cognitivist tradition. Whilst much categorisation research has been concerned with category representation - the encoding and structuring of objects together in some form of internal representation system (see Komatsu, 1992 for a review) - Wittgenstein was more concerned with word *use*, with the way that labels are used to pick out objects in the world as a part of the process of communication. In doing this, Wittgenstein was concerned with trying to specify the way in which the use of concepts and categories in communication imposes constraints on theoretical accounts regarding their nature - the 'looking' in §66 (quoted at the head of this chapter) - strongly emphasising the need to fully understand the problem before tackling any solution to it:

> 'We talk of processes and states, and leave their nature undecided. Sometime perhaps we will know more about them - we think. But that is just what commits us to a particular way of looking at the matter.' (Wittgenstein 1953, p102).

Wittgenstein was at pains in the *Philosophical Investigations* to start from as theoretically neutral a position as possible in his pursuit of an account of categorisation. In contrast, to start to examine categorisation from a standpoint which assumes stored conceptual representations, and processes of matching stimuli to them, is to start that investigation with an *a priori* commitment 'to a particular way of looking at the matter'. Thus, another factor that appears to elude cognitive science accounts of categorisation that attempt to incorporate Wittgenstein's theoretical standpoint is that in the *Philosophical Investigations* Wittgenstein doesn't just present some theoretical insights into human categorisation. Buried in his arguments is a thorough critique of the phenomena under examination - human concepts and categories, and the way words are attached to these - and the constraints that this critique imposes on any methodology for their study.

What follows is an attempt to elucidate those arguments in Wittgenstein's *Philosophical Investigations* relevant to the study of concepts and categories, and outline the constraints on the concept of concept that these investigations yield. The account developed will then be contrasted with the commonly held view of Wittgenstein's position in the cognitive science literature. I do not claim that the position I establish below is Wittgenstein's: as I acknowledged above, the questions asked by researchers in cognitive science are markedly different from those considered by Wittgenstein (for an excellent 'straight' exegesis of Wittgenstein's arguments see Baker and Hacker, 1980). Rather, my claim is: firstly that the position established here is consistent with the overall thrust of Wittgenstein's arguments; and secondly, that an analysis based on these arguments has much to offer contemporary debates regarding categorisation. Accordingly, having established a 'Wittgensteinian account of categorisation', in the following chapter I shall then evaluate the extent to which current theories of categorisation can accommodate the theoretical constraints established by this reading of Wittgenstein.

## 3.2 What Family Resemblances Are Not:

The accepted interpretation of Wittgenstein's account of concepts and categories in cognitive science is nicely summarised by Lakoff (1987a; though accounts which broadly concur with this can be found in Khatchadourian, 1966; Rosch and Mervis, 1975; Johnson-Laird, 1983; Holland, Holyoak, Nisbett and Thagard 1986; Komatsu, 1992).

Lakoff acknowledges Wittgenstein as the first theorist to notice what he terms a major crack in classical theories (e.g. Katz, 1972) of concepts and categories. These theories maintain that categories have clear boundaries and that they are defined in terms of common properties amongst a category's members, with these definitions perhaps being couched in terms of necessary and sufficient conditions for the determination of category membership (i.e. possession of properties X, Y and Z are both necessary and sufficient for an object to be considered an example of category N).

Lakoff notes that in PI §66, Wittgenstein argues that categories such as *game* cannot be accounted for according to the classical theory because there are no common properties that are shared by all games:

> 66. Consider for example the proceedings we call "games". I mean board-games, card-games, ball-games, Olympic-games, and so on. What is common to them all? - Don't say: "There *must*[2] be something common, or would they not be called 'games'" - but *look and see* whether there is anything common to *all*. - For, if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat, don't think, but look! - Look for example at board games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we next pass to ball games, much that is common is retained, but much is lost. - Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at a wall and catches it again, this feature has disappeared.

[2] The emphasis is Wittgenstein's.

> Look at the parts played by skill and luck; and at the differences between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristics have disappeared! And we can go through many, many other groups of games in the same way; can see how similarities crop up and disappear.
>
> And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.
>
> 67. I can think of no better expression to characterise these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. - And I shall say: 'games' form a family.
>
> And for instance, the kinds of number form a family in the same way. Why do we call something a "number"? Well perhaps because it has a - direct - relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.
>
> But if someone wished to say: "There is something common to all these constructions - namely the disjunction of all their common properties" - I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread - namely the continuous overlapping of those fibres". (Wittgenstein 1953, p31-2).

Lakoff draws two key theses from these passages:

*1: "Games, like family members are similar to one another in a variety of ways"; and

*2: "That [family resemblances], and not a single well defined collection of common properties is what makes game a category" (Lakoff, 1987a, pp 16-17)

From a close reading of §66 and §67, *1 would appear to be a fair statement of Wittgenstein's views. However, it is difficult to reconcile interpretation *2 with what

Wittgenstein actually says. In §66 Wittgenstein explicitly states that 'you will not see something that is common to all [games]'. Rather, he argues that what games have in common is the now notorious family resemblances:[3] 'a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail'. Lakoff, (and cognitive science literature in general) take this to be Wittgenstein's characterisation of what a category is. Consistent with Lakoff's position, Rosch and Mervis (1975) describe family resemblance relationships as consisting of a set of items in which

> each item has at least one, and probably several, elements in common with one or more other items, but no, or few, elements are common to all items... Members of a category come to be viewed as prototypical of the category as a whole in proportion to the extent to which they bear a family resemblance to (have attributes which overlap those of) other members of the category. Conversely, items viewed as most prototypical of one category will be those with least family resemblance to or membership in other categories. (Rosch and Mervis, 1975, p. 575)

In fairness, this description does appear to capture an important aspect of Wittgenstein's description of 'categories': however, what appears to escape Lakoff, Rosch, Mervis and other interpreters is the extreme negativity of this characterisation.[4] In PI §67 Wittgenstein explicitly condemns the vacuousness of accounts like this as a characterisation of concepts or categories. Saying that the common theme that runs through a category is the continual overlap of family resemblances is directly analogous to saying that the common thing that runs through a thread is continuous overlapping of the fibres that make up the thread, and Wittgenstein dismisses both of these accounts as empty gestures: 'Now you are only playing with words.' There is, he says, no thing that runs through a thread in the form of overlapping fibres, a thread simply is a series of overlapping fibres. Or to put this another way, as Baker and Hacker (1980) note, saying that the continuous overlapping of fibres is something

[3] 'Family resemblances' may well evoke something different in late twentieth century commentators than perhaps it did in Wittgenstein: it seems reasonable to assume that the product of a small nuclear family will have a different conception of family resemblance than someone from a complex, sprawling family unit whose numerous siblings embraced a wide range of diversities of age, taste, ability and even religion.

[4] "From the text it is obvious that the main thrust of this whole discussion is negative and critical" (Baker and Hacker, 1980 pp 327).

running through a thread makes the contrast between there being and there not being some thing running through the thread unintelligible.

Wittgenstein does not claim that family resemblances, and not a single well defined collection of common properties is what makes *game* a category; he attempts to show how empty this view is as a definition. His view here suggests a serious challenge to, rather than an endorsement of, Lakoff's formulation: if family resemblances are the common thing that run through game, just as overlapping fibres are the common thing that run through a thread, then what is this thing supposed to be?[5] How is it supposed to do whatever it is it is supposed to do? *How long,* Wittgenstein would appear to ask, *is a piece of string?*

## 3.3 The Length Of String - Categories And Boundaries

This question - 'how long is a piece of string?' - becomes important once the second part of Lakoff's exposition is introduced. Wittgenstein, as Lakoff notes, argues that the boundaries of categories are not fixed.

> 68. "All right: the concept of number is defined for you as the logical sum of these individual interrelated concepts: cardinal numbers, rational numbers, real numbers, etc.; and in the same way the concept of a game is the logical sum of a corresponding set of sub-concepts." - It need not be so. For I *can* give the concept 'number' rigid limits in this way, that is use the word "number" for a rigidly limited concept, but I can also use it so that the extension of the concept is *not* closed by a frontier. And this is how we do use the word "game". For how is the concept of a game bounded? What still counts as a game, and what no longer does? Can you give the boundary? No. You can *draw* one, for none has so far been drawn. (But that never troubled you when you used the word "game" before.)

[5] Goodman makes a similar point with regard to the explanatory power of unembellished "similarity": he argues that saying that two things are similar without stating the *respects* in which they are similar is entirely vacuous. 'To say all *a*'s are alike in being *a*'s amounts simply to saying that all *a*'s are *a*'s. The words "alike in being" add nothing; similarity becomes entirely superfluous.' (Goodman 1973 (p 439)).

> "But then the use of the word is unregulated, the 'game' we play with it is unregulated." - It is not everywhere circumscribed by rules; but no more than there are any rules for how high one throws the ball in tennis, or how hard; yet tennis is a game for all that and has rules too. (Wittgenstein 1953, p32-3).

Lakoff seizes upon the discussion of the category *number*. Historically, says Lakoff, numbers were first taken to be integers, and then 'numbers' were successively extended to include rational numbers, real numbers, complex numbers, transfinite numbers, and all of the other numbers that mathematicians are wont to invent. But the concept of 'number' is not bounded in any natural way, and it can be limited or extended depending upon one's circumstances and purposes. Lakoff says that in mathematics, intuitive human concepts like *number* must receive precise definitions: Wittgenstein's point, he claims, is that different mathematicians give different definitions, depending upon their goal. Thus although the category *number* can be given precise boundaries in many ways, 'the intuitive concept is not limited in any of those ways; rather, it is open to both limitations and extensions' (Lakoff, 1987a, pp 17).

The question therefore, on Lakoff's account, becomes one of how those limitations and extensions are governed - what factors determine the boundaries of categories in given circumstances. Lakoff answers this question in relation to *game* by saying that *game*'s boundaries are governed by resemblance to previous *games* in appropriate ways: a new thing can be a *game* if it is suitably similar to previous *games*. Lakoff gives the example of the introduction of video games in the 1970s as a recent case in history where the boundaries of the *game* category were extended on a large scale.

Once again, subtle and not-so subtle discrepancies can be distinguished between Lakoff's characterisation of Wittgenstein's views and the content of Wittgenstein's stated arguments. In §68, Wittgenstein says that one '*can* give the concept 'number' rigid limits in this way, that is use the word "number" for a rigidly limited concept,' - Lakoff's claim that in mathematics *number* must receive precise definitions appeals to this - 'but I can also use it so that the extension of the concept is *not* closed by a frontier.' Here, Wittgenstein is not talking about the extensibility of borders, but something far more radical: 'You can *draw* [a boundary], for none has so far been drawn. (But that never troubled you when you used the word "game" before.)'. Wittgenstein isn't talking here about the extensibility of boundaries; he is talking about their absence, a point developed in PI §69 to §73: categories do not have, or need, boundaries at all. In the context of Wittgenstein's overall discussion of categories, this is a vitally important point: it is one thing to seek to determine the length of a piece of string whose length isn't fixed (we might add a temporal dimension to our answer, for instance); it is quite another thing to seek to find out how long a piece of string is when the string is of no particular length at all.

> 69. How should we explain to someone what a game is? I imagine that we should describe *games* to him, and we might add: "This *and similar things* are called "games"". And do we know any more about it ourselves? Is it only other people that we cannot tell exactly what a game is? - But this is not ignorance. We do not know the boundaries because none have been drawn. To repeat, we can draw a boundary - for a special purpose. Does it take that to make the concept usable? Not at all! (Except for that special purpose.) No more than it took the definition 1 pace = 75cm. to make the measure of one pace usable. And if you want to say "But still, before that it wasn't an exact measure," then I reply: very well, it was an inexact one. - Though you still owe me a definition of exactness. (Wittgenstein 1953, p33).

In §69 Wittgenstein is emphatic. One *can* draw a boundary, for a special purpose, but it is just that, a drawn boundary. Important in the context of the special purpose, no doubt, but arbitrary to the concept or category in question. But we do not *need* to draw boundaries, because we can happily use concepts where none have been drawn; categories do not need boundaries to be usable. To further iterate this point, Wittgenstein considers the state of a user of a category (concept) who cannot specify that category's boundaries: is the user ignorant of those boundaries? - No, the user does not 'know the boundaries because none have been drawn'. Not knowing the boundaries of *game* is not a state of ignorance - it is reflective of the boundariless state of *game*.

§69, at least, seems to offer little support for the extensibility of boundaries as opposed to the lack of them; §70 serves only to reiterate and strengthen the boundariless case Wittgenstein makes in the *Philosophical Investigations* :

> 70. "But if the concept 'game' is uncircumscribed like that, you don't really know what you mean by a 'game'." - When I give the description: "The ground was covered with plants" - do you want to say that I don't

> know what I'm talking about until I can give a definition of plant?
>
> My meaning would be explained by, say, a drawing and the words "The ground looked exactly like this." - Then were just *this* grass and *these* leaves there, arranged just like this? No, that is not what it means. And I should not accept any picture as exact in *this* sense.
>
> Somebody says to me: show the children a game. I teach them to roll die for money, and the other person says to me "I didn't mean a game like that" Did he, when he gave me the command have to have in mind the exclusion of a game of dice? (Wittgenstein 1953, p33). [6]

This section thus underlines the boundariless point: one might draw a boundary (picture) in order to help understand a concept - to pin down a category - but this doesn't tell you exactly what was meant, since the concepts to be understood aren't fixed in that way. This point is crucial to Wittgenstein's argument: understanding concepts involves *conceptual* understanding; it does not concern some 'exact' understanding of states of affairs in the world (whatever such an understanding might be). At the end of §69, Wittgenstein commented that if one were to say to him that 'a pace' is not an exact measure, he would accept that it was an inexact measure, but he would also challenge his interlocutor for a definition of 'exactness'. In §70, he returns to this theme: even if he drew a picture of what he meant by 'the ground is covered with plants', and then said that what he was referring to 'looked *exactly* like this' he would not mean by its looking '*exactly* like this' that he meant that there was just *this* grass and *these* leaves, all arranged just like this. He would not accept any picture as being *exact* in this sense. This is because he isn't using - and we don't use - *exact* in that way. Exact is just another concept (categorising 'things that are exact'), which, whilst we can bound it in the special way, such as is envisaged by someone saying that 'a pace' is not an exact measure (defining exact in such and such a way), such a definition is not what *exact* is. One might say 'my desk is exactly five paces from the door', which would be another example of exact; one in which 'a pace' is an exact

[6] Lakoff suggests that this last paragraph should be read as indicating that Wittgenstein is suggesting that some category members are more central than others (see also Rosch, 1978), the idea being that dice is not a very good example (is atypical) of game. A more natural reading of Wittgenstein's intentions here is not that he wishes us to think of dice as a particularly atypical game, but rather that it is not an appropriate game to teach children, and that there is nothing in 'game' that rules this notion of appropriateness either in or out. (Thanks to Ulrike Hahn for the translation of this paragraph.)

measure (though in this case, 'exact' is not *used* in the same way, and with the same sense as by the interlocutor in respect of the plants in §70).

In §69 and §70 Wittgenstein makes use of this instinctive temptation to claim that boundariless categories are not exact, and therefore presumably problematical in use, to demonstrate the way in which we readily employ concepts - in this case *exact* - without boundaries unproblematically. One can make these points about *category* and *concept* as well. Wittgenstein's family resemblances 'definition' of these concepts is analogous to his discussion of *exact* above: a characterisation of how of *category* and *concept* are used (pre-theoretical 'concepts'), rather than some theoretical *definition* that states explicitly what these terms can stand for (i.e. as theoretical constructs). As Baker and Hacker (1980) observe, 'it is explanation, not definition, that is the correlate of understanding': being able to explain what a concept like game *is* is a criterion of understanding it (and as I shall show below, Wittgenstein is quite clear about what this does and does not entail); but someone's being unable to *define* a concept is not a criterion for determining that they do not understand it.

This thesis - that categories don't have boundaries - is vital to Wittgenstein's position, as witnessed by his working through another formulation of it in §71:

> 71. One might say that the concept 'game' is a concept with blurred edges. - "But is a blurred concept a concept at all?" - Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need?
>
> Frege compares a concept to an area and says that an area without boundaries cannot be called an area at all. This presumably means that we cannot do anything with it. - But is it senseless to say: "Stand roughly there"? Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand - as if I were indicating a particular *spot*. And this is just how one might explain to someone what a game is. One gives examples and intends them to be taken in a particular way. - I do not, however, mean by this he is supposed to see in those examples that common thing that I - for some reason - was unable to express; but that he is now going to *employ* those examples in a particular way. Here, giving examples is not an *indirect* means of explaining - in default of a better. For

> any general definition can be misunderstood too. The point is that *this* is how we play the game. (I mean the language game with the word "game".) (Wittgenstein 1953, p34).

Again, Wittgenstein's rejection of boundaries - and not just the idea of fixing upon this boundary rather than that one - seems to be both clear and unambiguous. We don't have to define boundaries in order to use concepts, nor is it clear that definite boundaries are always what we need; these points can be further drawn out if we contemplate §71 in conjunction with §76:

> 76. If someone were to draw a sharp boundary I could not acknowledge it as the one that I too always wanted to draw, or had drawn in my mind. For I did not want to draw one at all. His concept can be said to be not the same as mine, but akin to it. The kinship is that of two pictures, one of which consists of colour patches with vague contours, and the other of patches similarly shaped and distributed, but with clear contours. The kinship is just as undeniable as the difference. (Wittgenstein 1953, p36).

Categories - concepts - do not have boundaries, and by drawing boundaries we do not capture categories, we create something new - call them bounded categories (in §68 Wittgenstein calls them 'rigidly limited' concepts, so we might call our bounded *game* a rigidly limited game) - which have some kind of kinship with our natural categories, such as *game*, but a rigidly limited *game* is markedly and importantly different to *game*. (To return to a point made earlier, *using* concepts differs markedly from the pursuit of trying to form a theoretical view of them as constructs).

To return to family relations, these are the fibres that make up the threads that are categories; but Wittgenstein explicitly states that the length of these threads cannot be determined.

## 3.4 Categories, Essences And Schemas

Wittgenstein develops and expands the difference between the idea of an extensible boundary and the idea of no boundary throughout §72 to §83. However, yet another key thesis in Wittgenstein's account of categories was introduced in §71 above: in explaining what a game is, observes Wittgenstein, one gives examples of instances of games, and one intends those examples to be taken in a particular way. What one does *not* do, in giving these examples, is to expect the person to whom one is explaining 'game' to see the common thing - whether it be a core, schema or essence - which one cannot actually see oneself. It is true, says Wittgenstein, that when we give these examples our subject might see kinships between the examples, but these kinships are not in any way essential (to reverse the formulation of §76 - whilst applying the same logic - the *differences* between the instances will be just as undeniable as these kinships). Giving these examples, says Wittgenstein, is not an *indirect* explanation; it *is* the explanation. We *don't* give a general definition, but this is not because we can't think of one, but because there is none to give.

This theme is developed in turn in §72, §73, §74, §75 and §76.

> 72. *Seeing what is common.* Suppose I show someone various multi-coloured pictures, and say: "The colour you see in all these is called 'yellow ochre'". - This is a definition, and the other will get to understand it by looking for and seeing what is common to the pictures. Then he can look *at*, and point to, the common thing.
>
> Compare this with a case where I show him figures of different shapes all painted the same colour, and say: "What these have in common is called 'yellow ochre'".
>
> And compare this case: I show him samples of different shades of blue and say: "The colour that is common to all these is what I call 'blue'". (Wittgenstein 1953, p34).

Wittgenstein has already argued in §66, in respect of games, that if one looks at a category - that if one looks at games - one will not see any single thing that they have in common. Here, he questions the way that 'commonalities' are supposed to be garnered in the first place. In the first example, the commonality is easy to spot: provided the only common colour in the pictures was yellow ochre, and provided that the subject had grasped the meaning of colour, then she will be able to grasp what yellow ochre is - the colour that is common in all the pictures.

In example two, the subject could not proceed in the same way: although the figures all have colour (yellow ochre) in common, they also have other commonalities, such as being figures. Thus the subject could as easily learn to apply 'yellow ochre' to yellow ochre or to figures, or even to sample (all of the samples are samples after all) from this example. Nothing in the definition picks out the particular commonality that 'yellow ochre' is supposed to pick out.

Finally, in example three, there is no a priori colour commonality to the pictures; rather, the commonality can only be perceived if an observer *already* has the concept 'blue' (otherwise, she would see a riot of various 'colours'; not having any example of what 'not blue' is, she might also think that 'blue' just meant 'colours'). Since understanding this example is crucially dependent upon an understanding of 'blue', it follows that the example could not serve as a explanation of, or a definition of what is - or isn't -'blue'.

These points are then further expanded in §73:

> 73. When someone defines the names of colours for me by pointing to samples and saying "This colour is called 'blue', this 'green'..." this case can be compared in many respects to putting a table in my hands, with the words written under the colour samples. - Though this comparison may mislead in many ways. - One is now inclined to extend the comparison: to have understood the definition means to have in one's mind an idea of the thing defined, and that is a sample or a picture. So if I am shown various different leaves and told "This is called a 'leaf'", I get an idea of the shape of a leaf, a picture of it in my mind. - But what does that picture of a leaf look like when it does not show us any particular shape, but 'what is common to all shapes of leaf'? Which shade is the sample in my mind of the colour green - the sample of what is common to all shades of green.
>
> "But might there not be such 'general' samples? Say a schematic leaf, or a sample of *pure* green?" - Certainly there might. But for a schema to be understood as a *schema*, and not as the shape of a particular leaf, and for a slip of pure green to be understood as all that is greenish, and not as a sample of pure green - this in turn resides in the way that samples are used.
>
> Ask yourself, what *shape* must a sample of the colour green be? Should it be rectangular? Or would it then be the sample of a green rectangle? - So should it be 'irregular' in shape? And what is to prevent us from regarding it - that is, from using it - only as a sample of the irregularity of shape? (Wittgenstein 1953, p34-5).

Wittgenstein poses here a number of questions, albeit perhaps non-obvious ones, that are raised by the introduction of a generalised schema to serve as the basis for a category. (I should note that the objections Wittgenstein raises here apply to the idea of generalised category schemas in general, and not to some particular form of schema). Firstly, there is the question of the form that the generalisation should take; what shape should a generalised leaf be? Intricately linked to this is the question of how schemas are supposed to be used. Even if we can answer the first question - how we, say, generate a generalised temperature for ice-cream - we are still left with the related question of explaining how such a generalisation is used in practice. Which particular aspects of the schema are general, and which are not, and how in use are we supposed to know which is which. We might rephrase this question as asking which parts of the schema represent 'the generalised concept', and which are implementational details of the representation of this generalisation; if we relate this back to our earlier discussion of the A/B distinction, the question is how is the possessor of a concept schema to distinguish between those parts of the schema that are 'conceptual information', and those parts of the schema that are just aspects of the media for representing that information. Is the generalised green shape a schema for green or a schema for generalised shape? Which raises the further question: provided one could generate answers to these very challenging questions, what is supposed to be intrinsic to such a schema that would cause it to be used differently to an *example* of that which it was supposed to be a generalisation of? In §74, Wittgenstein explicitly claims that there is no such intrinsic feature in a schema, and thus that such a schema could not operate any differently from yet another sample of some thing:

> 74. Here also belongs the idea that if you see this leaf as a sample of 'leaf shape in general' you *see* it differently from someone who regards it as, say, a sample of this particular shape. Now this may well be so - though it is not so - for it would only be to say that, as a matter of experience, if you *see* the leaf in a particular way, you use it in such-and-such a way according to such-and-such rules. Of course there is such a thing as seeing *this* way or *that*; and there are also cases where whoever sees a sample like *this* will in general use it in *this* way, and whoever sees it otherwise in another way. For example, if you see the schematic drawing of a cube as a plane figure consisting of a square and two rhombi you will, perhaps, carry out the order "Bring me something like this" differently from someone who sees the picture three-dimensionally. (Wittgenstein 1953, p35).

Johnson-Laird (1983) offers a different interpretation of §73, claiming that it shows that Wittgenstein had adopted a Kant-like (1787) approach to schemas, arguing that 'a schema is not an image, but a model that underlies the ability to form an image' (Johnson-Laird, 1983, p 190). Seizing on Wittgenstein's comment that 'for a schema to be understood as a *schema* ... in turn resides in the way that samples are used' Johnson-Laird claims that Wittgenstein is arguing that schemas do act as representative samples 'in the same way as mental models acts as representative samples' (Johnson-Laird, 1983, p 190): according to Johnson-Laird, Wittgenstein argues in this section for a view of schemas that are '*not* a set of necessary and sufficient conditions, but the typical or default characteristics of the items [they] subsume.'

As with the standard view of family resemblances, this interpretation sits ill in the context of the broader argument advanced in these passages. Wittgenstein introduces the idea of schemas here not because he advocates them as a basis for the mental representation of concepts, but rather in order to show how despite possessing a certain theoretical seductiveness, schemas cannot provide answers to questions about the mental representation of concepts: 'this may well be so - though it is not so'.

As I noted earlier Wittgenstein's position in objecting to the idea of schemas as a basis for concepts is far more systematic than Johnson-Laird's isolated example indicates. Wittgenstein details a list of requirements that schemas would have to fulfil if they are to act as a basis for categories:

- What form is a schema supposed to take: what shape is a generalised leaf?;
- Which aspects of a schema are generic, and which are not, and how do we explain the distinction between generic and non-generic aspects of the representation in the use of the proposed schema?
- What is supposed to be intrinsic to a schema that would cause it to be used differently to examples of whatever it was that it was supposed to be a generalisation (or model) of?

Wittgenstein makes quite clear in §71 - 74 that he feels that there simply are no satisfactory answers to these questions. Wittgenstein, in these passages, is quite clearly not advocating schemas as a theory of category representation; rather he is seeking to demonstrate - convincingly - that schemas alone *cannot* provide an account of how concepts are represented.[7]

## 3.5 Samples And Similarities: Beyond The Boundary

So far the general focus of this discussion has been on what Wittgenstein says categories, and by extension categorisation, are not. In doing this, however, I have introduced some aspects of Wittgenstein's account of what it is to 'know' a category - what having a concept of something is - in the discussion of the sections from the *Philosophical Investigations* above. In §69, Wittgenstein remarked 'How should we explain to someone what a game is? I imagine that we should describe *games* to him, and we might add: "This *and similar things* are called "games"". And do we know any more about it ourselves?'; in §71 '[how might one] explain to someone what a game is[?] One gives examples and intends them to be taken in a particular way. - I do not, however, mean by this he is supposed to see in those examples that common thing that I - for some reason - was unable to express; but that he is now going to *employ* those examples in a particular way'. Wittgenstein argues that we learn *game* from examples of games, and that our understanding of *game* is dependent upon how we re-employ these examples.

This is one of the ways in which Wittgenstein's thesis that 'meaning is use' is introduced in the *Philosophical Investigations*. It should be noted, however, that '*meaning is use*' does not express the rather banal truism that meanings are determined by how we use concepts (or words), but rather it expresses the more fundamental idea that there simply *is no more* to meaning than how the examples which comprise a category are used. This is clear as even in the brief remark above: the concept of *game* is no more and no less than the way 'game' and examples of games are used.

[7] Some similar points are made by Winograd and Flores (1986) with respect to the use of simple frames to represent concepts. With respect to frames, they note 'if we look at the literature on frame systems [for an answer to such questions] we find a mixture of hand-waving and silence. Simple rules don't work. If, for example, defaults are used precisely when there is no explicit (previously derived) information to the contrary, then we will assume that one holds even when a simple straightforward deduction might contradict it. If analogies are treated too simply, we attempt to carry over detailed properties of one object to another for which they are not appropriate (p117)'. There is a strong parallel to be drawn between the way frames are used in artificial intelligence, and the role schemas are supposed to play in cognitive theories of categorisation.

> 75. What does it mean to know what a game is? What does it mean, to know it and not be able to say it? Is this knowledge somehow equivalent to an unformulated definition? So if it were formulated I should be able to recognise it as the expression of my knowledge? Isn't my knowledge, my concept of a game, completely expressed in the explanations I could give? That is, in my describing examples of various kinds of game; showing how all sorts of other games can be constructed on the analogy of these; saying that I should scarcely include this or this among games; and so on. (Wittgenstein 1953, p35).

There are no boundaries, common essences or anything else says Wittgenstein: the 'common things' that we - for some reason - are unable to express. Rather, categories simply comprise samples (connected by the network of commonalities alluded to in 'family resemblances') and rules governing the use of these samples (use is made possible by our grasping that the samples are to be 'taken in a particular way').

Rundle (1996) claims (taking §75 in isolation) that Wittgenstein uses this point to establish the idea that definitions are not something we *consult*, but that his arguments do not 'exclude definitions from another role, namely that of specifying the various features of things we look to in making our classifications' (pp 67). Rundle takes the difference in meaning between 'bottle' and 'jar' as an example: 'although this question may never have engaged us explicitly, it takes very little reflection to conclude that we use 'jar' when the width of a container's opening is close in size to the body of the container, whereas with a bottle we have a neck which tapers to a relatively narrower opening... it is clear that it is width of opening (amongst other things) that guides us in the use of these terms' (pp 66-7). Not only is this interpretation clearly at odds with Wittgenstein's arguments in the sections foregoing this passage - Wittgenstein is at pains to demonstrate that *game* cannot be given any such definition - but as studies by Sloman Malt, Shi, Gennari and Wang (1997a; 1997b) and Labov (1973) demonstrate, it is also clearly at odds with empirical data regarding the way people actually use the terms 'bottle' and 'jar'. Labov found that drawing the kind of distinction Rundle envisages between membership and non-membership for simple physical object categories such as *cup* and *bowl* was less than straightforward; the differences between *cup* and *bowl* vary along a continuum, and different participants put the cut off point on the continuum in different places. Moreover, this point could be altered by context: if participants were asked to imagine an object which is otherwise half-way between a *cup* and a *bowl* as containing mashed potato, then participants showed a marked preference for considering the object to be a *bowl.*

Further evidence is provided by Sloman et al, who found in their study that there was a clear distinction between naming and the kind of perceptual similarity Rundle invokes; names used to describe objects in their study did not appear to be governed by perceptual similarity relations and that names did not "exert a strong influence on the shape on conceptual similarity space, or on object locations in that space. Learning to call [a] blue smurf a "juice box" may make that object a member of the linguistic category "box" without making it seem more similar to rectangular things made of cardboard." (Sloman et al, 1997b, p239). [8]

The key point Wittgenstein is seeking to establish here is not that one be able to specify the relevant commonalities, - 'What does it mean to know what a game is? What does it mean, to know it and not be able to say it?' - since it might be, as with *game*, that one cannot. Rather, what is important is the way commonalities are considered: knowing 'examples of various kinds of game' and also 'how all sorts of other games can be constructed on the analogy of these'. Wittgenstein is differentiating here between that which we can *specify*, the examples of games (or bottles and jars) that we have encountered previously, and the underlying process by which these previously encountered samples can inform both our understanding of new samples, and of the terms *game*- or *bottle*, or *jar* - in general.

Wittgenstein is also making the point that knowing how we make these considerations - our being able to specify or describe the underlying cognitive processes - is *not* relevant to our understanding of *game* (just as being able to specify our capacity for language is irrelevant to conversation). This is further illustrated in much of the remainder of the *Philosophical Investigations*, where Wittgenstein's concern for underlying processes - such as saying how samples are to be taken in such and such a way - is a key motivation for the considerable amount of time devoted to Wittgenstein's extensive working through of the idea of 'rule following,' where Wittgenstein takes great care to show that following rules is a cognitive process that

8 Sloman et al go on to say "... In that sense, it [the blue smurf] may not be thought of as a box even though it is called a box" (p 239): I note this, since it seems firstly to beg the question of what a box is - in Wittgensteinian terms one might say: 'Alright, but now you owe me a definition of "box"' - and also because it begs the further question of how one might 'think about' something that one calls a "box" without thinking of it as a box!

most definitely isn't conscious or deliberate, but is rather instinctive and innate (c.f. McGinn, 1983; Baker and Hacker, 1983).

## 3.6 Representation And Process

As I noted above, in §76, Wittgenstein argues that 'if someone were to draw a sharp boundary' around a concept, he could not acknowledge that concept as being identical to the concept that he possessed whose name corresponded to it; and that the differences between such a bounded concept and a 'natural concept' would be as significant as any similarities between them. Taken together, §76, 77 and 78 develop an argument that firmly opposes the idea of definitions (and fixed category representations) altogether:

> 76. If someone were to draw a sharp boundary I could not acknowledge it as the one that I too always wanted to draw, or had drawn in my mind. For I did not want to draw one at all. His concept can be said to be not the same as mine, but akin to it. The kinship is that of two pictures, one of which consists of colour patches with vague contours, and the other of patches similarly shaped and distributed, but with clear contours. The kinship is just as undeniable as the difference.
>
> 77. And if we carry this comparison still further it is clear that the degree to which the sharp picture *can* resemble the blurred one depends upon the latter's degree of vagueness. For imagine having to sketch a sharply defined picture 'corresponding' to a blurred one. In the latter there is a blurred red rectangle: for it you put down a sharply defined one. Of course - several such sharply defined rectangles can be drawn to correspond to the indefinite one. - But if the colours in the original merge without a hint of any outline won't it become a hopeless task to draw a sharp picture corresponding to the blurred one? Won't you then have to say: "Here I may as well draw a circle or a heart as a rectangle, for all the colours merge. Anything - and nothing - is right." - And this is the position you are in if you look for definitions corresponding to our concepts in aesthetics or ethics.
>
> In such a difficulty, always ask yourself: How did we *learn* the meaning of this word ("good" for instance)? From what sort of examples?

In what language games? Then it will be easier for you to see that the word must have a family of meanings.

78. Compare *knowing* and *saying*:

how many feet high Mont Blanc is-
how the word "game" is used-
how a clarinet sounds.

If you are surprised that one can know something and not be able to say it, you are perhaps thinking of a case like the first. Certainly not of one like the third. (Wittgenstein 1953, p36).

These sections underline the arguments developed so far: not only can we not bound category space, but any attempts to do so will be pointless, necessarily unsuccessful. One cannot use a bounded 'definition' or 'representation' to characterise an unbounded phenomenon. A sharp picture cannot characterise a blurred picture; they are simply two different pictures. Concepts are learned by the examples that typify them, and Wittgenstein has already argued that no common thread between these examples can define a given concept or category. Not only will any definition necessarily not capture the concept, but there is no end to the variety of definitions which one could equally validly (or rather invalidly) put forward; 'I may as well draw a circle or a heart as a rectangle, for all the colours merge.'

It seems reasonable to conclude from this that Wittgenstein is not arguing for a treatment of concepts and categories, but rather that he is trying to undermine our pre-theoretical attachment to a particular idea of what concepts and categories are in order to demonstrate that in asking what a concept or category is in this way - *with a particular determinate answer in mind* - we are asking the wrong question. Asking what a concept or category is is like asking how a clarinet sounds, not like enquiring after the height of mountains (that is to say, it is like asking the height of mountains in general, and not the height of a particular mountain).

This is a vitally important, and often overlooked, aspect of Wittgenstein's treatment of categories: namely that to Wittgenstein, if we ask the question 'what makes a category a category?' we should not expect to find an answer, not because the answer is somehow 'beyond our ken', but because we are asking - fundamentally - the wrong question. Elsewhere in the Philosophical Investigations (p. 308) Wittgenstein examines the way in which problems about mental processes[9] and states arise. He argues that the most important step in the production of such problems is the one that goes unnoticed - the first one:

> 'We talk of process and states, and leave their nature undecided. Sometime perhaps we will know more about them - we think. But that is just what commits us to a particular way of looking at the matter.' (Wittgenstein 1953, p102).

Wittgenstein's arguments, as examined so far, do not advocate a particular view of categories - what has become known loosely as 'family resemblance theory' - but rather they represent a thorough attempt to eliminate the temptation to see the question of how humans categorise things - how it is this or that 'thing' are considered to be, say, 'games' - in terms of determinate, or even necessarily determinable, 'categories'. To Wittgenstein, the problems involved in explaining how categories are defined stem not from the phenomenon under examination, but the way this phenomenon has traditionally been defined; hence, perhaps, the famous 'don't think, but look!' (famously criticised as being an inadequate basis for studying categorisation by Medin and Ortony, 1989). If we 'think', i.e. if we assume that the existence of things called games entails the existence of categories - defined in some as yet to be determined way - in virtue of which the things can be considered games, we are not exploring categorisation: we are predetermining what it can be. And since categorisation does not conform to the picture we have attempted to impose upon it, we find that our subsequent attempts at exploration are difficult and frustrating (one might draw an analogy between this process, and that of exploring a lost continent with a map one had drawn before one set out). However, argues Wittgenstein, if we 'look', we will find that categorisation does not conform to our pre-theoretical expectations: the existence of things called games *does not* entail the existence of 'categories' if what we mean by category is some determinate *thing* in *virtue of which* those things can be considered games. For if we look, we see that the use of concepts such as 'game' precludes the existence of simple category representations in virtue of which such things can be considered games, since 'game' is used in such a way that there is nothing *in particular* in virtue of which 'games' are games. Rather, there are a number

[9] As I noted earlier, Wittgenstein actually considers the way *philosophical* problems concerning mental processes arise - his analysis does, however seem equally applicable to psychological problems and questions.

of things - samples - in virtue of which 'games' are games - '[a] word must have a family of meanings' - and some process - the 'appropriate ways' of taking these samples - by which they contribute to the category (or categories) *game*.

## 3.7 Discussion

The broad outline of Wittgenstein's argument can be summarised as follows:

1. That categories have no necessary or sufficient defining characteristics: rather that kinships - "family resemblances" - can be traced across categories (§65-7);

2. That these category spaces are unbounded - i.e. there are no boundaries to the space across which "family resemblances" can be traced (§68, 69, 70, 71, 73);

3. That learning a category such as game does not involve extracting an essence or schema from instances. (§71-83)

4. In learning a "category" such as game, one learns examples (instances) and appropriate ways of using these examples (§69,71, 73, 81, 82)

Taken together, these arguments don't amount to a quibble about the constitution of categories; rather, I have sought here to show that Wittgenstein was rejecting altogether the idea of categories or concepts as definitive, determinable, theoretical constructs. He seeks to demonstrate in the *Philosophical Investigations* that the idea that there is a *thing* (some specific representation, even when called a category) that can determine whether a game is a game makes as much sense as the idea that there is some *thing* that runs through a whole thread - namely the continuous overlapping of those fibres; and this is an idea that Wittgenstein rejects as a vacuous play on words. What is a category, a game? How long is a piece of string? It isn't that the answer to these questions is hard to find, but rather that the questions make no sense (are nonsense).

Another way of seeing this same point is to view Wittgenstein's argument as a challenge that any strong theory of category representation - i.e. a theory couched in purely representational terms - would have to meet. In order to explain what a category is, such a theory would have to explain how a category essence or schema can be a determinant of categories which have no necessary or sufficient defining characteristics, but rather kinships of "family resemblances" that can be traced across them, and whose category spaces are unbounded - i.e. there are no boundaries to the space across which "family resemblances" can be traced. It would appear that Wittgenstein felt - with very good reason - that these are challenges that cannot be met.

In the following chapter, I shall examine the empirical evidence that has been collected regarding categorisation and the current state of cognitive models of categorisation. I shall attempt to show that Wittgenstein's critique, as presented here, is consistent with the empirical evidence[10], and then explore the implications this has for analogy and categorisation.

[10] Moreover, I shall argue Wittgenstein's treatment of 'concepts' can provide the theoretical glue to bind the mass of psychological evidence together.

# Chapter 4

# Categorisation II
# Theories and Models of Categorisation

## *4.1 The Challenge*

What emerged from the foregoing examination of Wittgenstein's theoretical analysis is not any kind of working definition of categorisation, but rather a fundamental challenge to our common concept of concepts. Wittgenstein's arguments throw into question the very idea that a given category embraces a determinate, identifiable concept.

If Wittgenstein is right, then the consequences for a view of cognition that embraces distinct, separable processes of categorisation and analogy could be far-reaching. As I argued in chapter 2, in cognitive science analogies have traditionally been defined as being distinct - somehow - from categories, the nature of which are left unexamined with respect to such definitions (although it has been assumed that concepts and categories are somehow specified and represented in a way analogies are not).

I mentioned earlier that without an account of categorisation couched in terms of specifically - and discernibly - represented concepts and categories, a distinction between analogy and categorisation which is reliant upon a contrast of analogy with categorisation won't actually do any distinguishing at all. Without such an account of categorisation, a typical definition of analogy such as the one mentioned in chapter 2 :

> *"In an analogy, a familiar domain is used to understand a novel domain in order to highlight important similarities between the domains, or to predict new features of the novel domain"*
> (Clement and Gentner, 1991, p 89).

would be equivalent to:

> *"In PROCESS, a stored representation is used in order to highlight important similarities between it and a new representation of an object or concept, or to predict new features in the new representation of an object or concept."*

where *PROCESS* could equally validly be replaced by analogy or categorisation, since there would be no way of distinguishing between the two.

In this chapter,[1] I shall examine current knowledge and perspectives regarding categorisation in the light of Wittgenstein's arguments. Does the empirical evidence match Wittgenstein's analysis? Can that analysis shed some light on the mass of existing data? Or are there empirical findings that offer up the possibility of our retaining a determinate, theoretical view of concepts, thereby salvaging the traditional analogy - categorisation distinction?

## 4.2 Definitions And Boundaries

Wittgenstein's first two sets of arguments attacked variations on what is usually referred to as the definitional or 'classical' view of concepts (see Smith and Medin, 1981), which holds that concepts possess definitions specifying features *necessary* and *sufficient* for the concept.

Wittgenstein argued:

1. Categories have no necessary or sufficient defining characteristics: rather that kinships - "family resemblances" - can be traced across particular categories (§65-7); and
2. The spaces these categories cover are unbounded - i.e. there are no boundaries to the space across which "family resemblances" can be traced (§68, 69, 70, 71, 73);

---

[1]I am indebted to Ulrike Hahn for the many discussions that have influenced many of the ideas articulated here. In particular, two review papers Dr. Hahn and I wrote collaboratively have helped determine the shape of this chapter. I have no wish to claim sole authorship over what is very much joint work, and hence, for the purposes of this thesis, I shall reference these papers as I would the work of any other outside author.

As these two arguments are jointly aimed at an earlier view of concepts put forward by Frege (1952; a variant had been previously held by Wittgenstein in the *Tractatus*, 1922) they are best considered jointly.

Frege characterised concepts in terms of sets of defining attributes, distinguishing between a concept's *intension* and its *extension*. In the psychological literature, this distinction has generally been interpreted and expressed as follows: the intension of a concept is the set of attributes that define what it is to be a member of the concept, and the extension is the set of items that are members of the concept. Thus, on this basis, a concept's feature-definition is the summary description of the entire class used in every instance of categorisation, which proceeds simply by checking for the presence of these features in the instance in question, and a concept's class-definition is a list of all those instances that have been encountered that possess these features. If we take the concept *bachelor*, its intension would include characteristics such as *male, unmarried* and *adult*, and its extension would be the set of male, unmarried, adults in the world. This *defining-characteristics* view is often supplemented by a "nesting assumption": that a subordinate concept (e.g. *robin*) contains nested within in it the defining features of the super-ordinate (*bird*); see for example Collins and Quillian (1969; 1970).

Thus, according to the *definitional* view, concepts are defined in terms of a set of conjunctive attributes, each of which is necessary, and all of which are jointly sufficient for an item to be recognised as being an instance - or member - of that concept or category. As what is and what is not a member of a category is therefore clearly defined, clear-cut boundaries between what counts as a member and non-member of a category can be posited.

### 4.2.1 Definitions

As Ramscar and Hahn (1998a) argue, despite some evidence that definitions can form the basis for categories in empirical studies using artificial stimuli (e.g. Bruner, Goodnow and Austin, 1956), Wittgenstein's analysis of the definitional view is confirmed when studies transfer their focus from artificial concepts in controlled experiments to our everyday concepts (i.e. the concepts for which we typically have words). When it comes to typical, everyday concepts, definitions appear inadequate, as do theories that are based upon them.

Ramscar and Hahn observe that of the difficulties the definitional view encounters with everyday concepts, the most serious is the one highlighted by Wittgenstein in regard to games: almost all everyday concepts appear to be indefinable (Fodor, Garrett, Walker and Parkes, 1980; Fodor 1981). It simply hasn't been possible to formulate necessary and sufficient conditions for a given item's being a game; or a chair, or a bowl, or a smile. Moreover, if some form of the definitional view is correct, it seems reasonable to assume that people should, if asked, be able to name the features for all the members of a concept, and generate fairly similar features sets that are in some way congruent with a definitional feature-set for that concept. Yet studies by Conrad (1972) and Rosch and Mervis (1975) have shown that this is not the case; participants list a range of non-necessary attributes instead. Rosch (1973) found that participants rated the typicality of category members differently (e.g. *robin* is a more typical bird than *canary*), and that this typicality could be used as a predictor in the times participants took to confirm statements (e.g. "a robin is a bird"); these results again sit ill with the apparent democracy of a defining-characteristic definition of a concept.

These problems with definitions are further illustrated by Ramscar and Hahn's observation that common dictionary "definitions" of almost all terms are not really definitions at all, in that they do not provide necessary and sufficient conditions for category membership. Instead they typically do no more than provide some relevant information about category members in order to help the dictionary user attempt to identify which concept is intended (Ramscar and Hahn, 1998a).

### 4.2.2 Boundaries

Further evidence contra the definitional view comes from examining the boundaries of categories. The definitional view implies that these are distinct, and that they should cleanly separate instances from non-instances. However, consistent with Wittgenstein's arguments, this turns out not to be the case.

Theoretically, there are many other arguments against category boundaries that can be laid alongside Wittgenstein's. Ramscar and Hahn (1998a) use a thought experiment originally put forward by Black (1949) to illustrate how determining a category's boundary almost always involves a degree of vagueness. If one is asked to imagine a series of 'chairs' differing in quality by least noticeable amounts, one can end up imagining something like an ordered sequence which moves from a Chippendale chair on the one end to a lump of wood at the other end. A 'normal' observer should find it extremely difficult to point to the dividing line between 'chairs' and 'non-chairs' along this continuum[2]. As Black makes clear, this uncertainty over category boundaries can be generated for *any* term whose application requires the use of a 'sense', or bounded meaning; that is to say, such uncertainties apply to *all* 'material' terms.

Another factor that casts doubt on the idea of clearly specifiable boundaries is noted by Lakoff (1987b). Even when concepts do appear to have definitions, these definitions generally hold only with respect to a range of 'background assumptions'. Varying these assumptions immediately produces unclear or borderline cases:

> "The noun bachelor can be defined as an unmarried adult man, but the noun clearly exists as a motivated device for categorizing people only in the context of a human society in which certain expectations about marriage and marriageable age obtain. Male participants in long-term unmarried couplings would not ordinarily be described as bachelors; a boy abandoned in the jungle and grown to maturity away from contact with human society would not be called a bachelor." (Fillmore (1982), quoted in Lakoff 1987b)

Background factors, such as the social conventions concerning marriage, will, in general, hold to varying degrees, such that, the argument goes, the definition of bachelor can meaningfully be applied if background conditions are sufficiently similar to the conventions concerning marriage current in the West. But even then, to take as an illustration the particular marital circumstances of the Pope, it is still far from clear that the simple definition 'unmarried male' is sufficient to explain *all* of the correct and incorrect applications of the concept *bachelor.*

Alongside these further theoretical analyses, Ramscar and Hahn point to compelling empirical evidence that suggests that many - if not all - natural language categories do not have clear boundaries. Studies such as the one noted in chapter 3 by Labov (1973; 1978) which focused on cup-like containers, examining the variability that was inherent in the use of terms such as *cup*, *bowl*, *mug* etc., between participants and between contexts lends empirical support to the boundariless analysis. Labov found that there was no clear boundary between membership and non-membership for simple

[2]The difficulties posed by continua are old, as illustrated by the *sorites* (heap) and *phalakro* (bald man) paradoxes, which originate with the Megarian philosophers in the early 4th century BC, (Barnes, 1979)

physical object categories such as *cup* and *bowl*, but rather, confirming Black's thought experiment, participants responses indicated that the differences between *cup* and *bowl* vary along a continuum, with different participants putting the cut off point on the continuum in different places. Moreover, this point could be altered by context: if participants were asked to imagine an object which is otherwise half-way between a *cup* and a *bowl* as containing mashed potato, then participants showed a marked preference for considering the object to be a *bowl*. However, as Ramscar and Hahn, (1998a) note, Labov's interest was primarily in formalising the variability found, and hence his results are not presented with the detail experimental psychologists might want. These results receive more formal support from those of McCloskey and Glucksberg (1978) who presented a study of 540 exemplar-category pairs (e.g., *apple-fruit*) which revealed not only substantial between and within participant disagreement over category membership (the latter measured over successive test-sessions) but also showed levels of disagreement to correlate with independently derived typicality ratings. For example, McCloskey and Glucksberg's participants were certain that *chair* belonged to the category *furniture*, and that *cucumber* did not. However, McCloskey and Glucksberg found that there was much between-participant disagreement as to whether *book-ends* belonged to the category *furniture*, and even that individual participants differed in their judgements from one session to another.

### 4.2.3 The Case For Definitions And Boundaries

Consistent with Wittgenstein's first two sets of arguments, what evidence there is in the literature does tend to confirm the view that categories have no necessary or sufficient defining characteristics; rather some as yet unspecified kinships - *family resemblances* - can be traced across categories. Moreover, there is also confirmation that the boundaries of the spaces these categories cover cannot be specified - i.e. there are no clearly bounded spaces across which "family resemblances" can be traced. However, it is worth noting (perhaps with some amazement) at this juncture that despite the oft-made claim regarding the influence of Wittgenstein on the study of concepts and categories, and the extensive empirical activity in this area in the past 25 years, very few studies have actually directly addressed these two points (which as I argued above, are vital elements in the theoretical challenge Wittgenstein presents).

## 4.3 Defining Essences

Wittgenstein's final two sets of arguments reject the idea of some abstracted schema in preference for an account based on previously encountered examples. As I noted in Chapter 3, he claims that:

3. Learning a category such as game does not involve learning or extracting any essence or schema from the instances (§71-83);
4. In learning a "category" such as game, one learns examples (instances) and appropriate ways of using these examples (§69,72,73,74,81,82).

Whilst, as Komatsu (1992) notes, the vast majority of experimental results do not directly indicate *anything* about conceptual representation: separating form, content and the processes acting on concepts is an invidious business (best illustrated by Wittgenstein's remarks on schemas above) the *issue* of whether or not a particular learning process involves the abstraction of some core essence - be it a schema, a theory or a prototype - or not has been central to experimental psychology. Because a number of excellent reviews of theories of categorisation and conceptual structure already exist (Smith and Medin, 1981; Komatsu, 1992), I propose, in the following two sections, to eschew exhaustiveness, and will not attempt to present an exegesis of every theory of categorisation that has ever been presented. Instead, I shall focus on the theories with widest common currency: the *prototype* theory (Rosch and Mervis, 1975; Rosch, 1978; Hampton, 1979); the *exemplar* theory (e.g. Medin and Schaffer, 1978; Nosofsky, 1986, 1987, 1988, 1989, 1991, Nosofsky, Clark and Shin, 1989, Shin and Nosofsky, 1992; Kruschke, 1992; Lamberts, 1996); and the *theory (or explanation-based)* theory (Carey, 1985; Medin and Murphy, 1985; Gelman and Markman, 1986; Keil, 1989; Medin and Ortony, 1989; Gopnik and Meltzoff, 1997). The *exemplar* theory - which loosely, states that category membership is computed according to a comparison with known category members - is compatible with the account of Wittgenstein's views presented above, whereas the *prototype* theory and the 'theory' theory[3] are compatible with the contrary notion of determinate, specifiable concepts.

[3] Which has been proposed to an extent - c.f. Medin and Ortony (1989) - in response to a dissatisfaction with both the'traditional' Wittgensteinian view, and prototype and exemplar accounts.

## 4.4 Prototypes As Essences

"Prototype" - as a theoretical term - has been used loosely. As I noted in chapter 3, the prototype - or family resemblance - account of categorisation has usually (and, in the light of the foregoing discussion, erroneously) traced its roots back to Wittgenstein's 'account' of categorisation in the *Philosophical Investigations.* In this review, I shall use prototype to refer exclusively to those theories that equate category representation with the storage of central tendencies for categories.

The prototype account is probabilistic, in the sense that most of the features (or characteristics) associated with a concept are in some way *likely* to be found in instances, but this likelihood is not absolute. Instead, a *family resemblance* structure is proposed: instances of a category are thought to resemble one another in the way that members of a family do (Rosch and Mervis, 1975). Thus, according to this view, some members of a family will be very *typical*, as they will share many features (probabilisitically) with many other family members, whereas some family members may be very *atypical* because they share only a few features with other family members.

The features that are central to prototype accounts of categories can be summarised as following:

1. *Centrality of typicality*: the degree of typicality is the central conceptual mechanism; understanding a concept is understanding typicality, and the probabilities related to it.

2. *Abstractness*: each attribute specified for a concept is shared by more than one instance of a concept; thus by abstraction, category coherence is governed by an overlapping network of shared attributes.

3. *Weighted attributes*: attributes are 'weighted' according to their relevance to judgements of category inclusion; when a sufficient weighting of attributes is possessed by an instance (which thereby exceeds some threshold), it will be considered to be a member of a given category (Rosch and Mervis, 1975). The more weighted attributes an instance possesses, the more typical it will be judged to be.

4. *Combinational weighting*: concepts are defined by combinations of attribute weights. Attribute weights are independent, but are combined through adding (Rosch and Mervis, 1975; Smith and Medin, 1981); Komatsu (1992) notes that in consequence, as implied by the thresholds described in point 3, instances and non-instances of a concept can be neatly partitioned using a linear discriminant function. If one has to plot a set of objects by the combined weights of all their attributes, all instances would fall on one side of a category's boundary line (with all noninstances falling on the other side of that line).

5. *Retention of central tendencies*: the corollary of the foregoing, therefore, is that knowing a concept involves being in possession of summary knowledge of the central tendencies of instances of a given category, and hence, *not* being in possession of representations of individual instances.

### 4.4.1 Prototypes, Family Resemblances - And Wittgenstein

The gulf between prototype accounts of categorisation on one hand, and Wittgenstein's theoretical account of the nature of family resemblances on the other, is, I hope, readily apparent even before I formally contrast the two. Prototype theories are, in essence, attempts to enumerate and formalise an account of categorisation that Wittgenstein explicitly condemned as vacuous - 'now you are just playing with words' - in the *Philosophical Investigations.*

From the point of view of this critique, I shall argue the chief failing of prototype theories of categorisation is that they violate Wittgenstein's criteria for determining what knowing a concept entails, whilst simultaneously failing to meet the important challenges that are implicit in these criteria.

#### 4.4.1.1 Boundaries

Firstly, prototype theories are unable to meet Wittgenstein's challenges regarding definitions of boundaries. As Komatsu (1992) argues, without an account of what is required to determine a priori constraints upon the level of - or the nature of - the similarities that are shared by category instances, it is unclear exactly what principled account can be provided by prototype theories of *which* similarities are to count in deciding the boundaries between concepts, and *how* these are to apply. Take, as an example, a killer whale and a mouse. Killer whales and mice do not appear to share a vast quantity of similarities, but presumably they share enough for both to be considered mammals. However, a killer whale also appears to share as many similarities with a shark, which is a fish. Without a principled account of how similarities are to be determined to be particularly category salient, this tends in turn to beg the question of why a killer whale is a mammal and not a fish? (Or at least, why killer whales and sharks belong to different categories.)

Komatsu notes that presumably a prototype account would predict that the summed weights of killer whale attributes lead to its being more similar to other mammals than fish, and hence a killer whale is categorised as a mammal. However, in determining *these* weights, we need first to know how common these attributes are to both mammals and fish, which necessitates the categorisation of mammals and fish prior to the categorisation of killer whales. Yet presumably the determining of weightings in mammals involves a computation of the similarities across all mammals - including killer whales - which begs the question, given the similarities between killer whales and fish, of why a particular boundary was drawn - between killer whales and similar fish - in the computation of those weightings. To put this in a form that makes the circularity more explicit: a prototype-style categorisation of killer whales as mammals is contingent upon the categorisation of mammals and fish, but categorisation of mammals and fish assumes a prior categorical separation between killer whales and similar fish, which is dependant upon how both in turn are categorised...!

One attempt to avoid this apparent circularity has been to argue that certain partitionings of the world (including, one would assume, mammals and fish) are privileged, in that they arise naturally from the interactions of our perceptual apparatus and the environment, and are therefore more immediate or direct (Rosch, 1978; Anderson, 1990). According to this argument, similarity weightings are constrained ecologically, reflecting a natural partitioning in the world that is imposed by our perceptual systems.

I noted earlier Wittgenstein's famous observation that any talk of "processes" and "states of affairs" in the world will entail an a priori commitment to a particular way of looking at matters (Wittgenstein 1953, p102). Choosing what is to count as facts of the matter - which categories are 'ecological' - when it comes to categorisation will act as a powerful a priori determinant of the picture of the process one will uncover. Thus taking on board a different set of a priori facts will radically alter this picture. The argument from ecology described above ignores any consideration of the idea that all classification systems are ultimately human constructs (indeed, it is an example of letting one's immersion in one such system blind one to any alternatives).

Consider the following example from 'Ontology', (*Moby Dick,* Melville, 1851), where the central character, Ishmael, examines all of the reasons put forward by Linnaeus for classifying whales as mammals:

> First: The uncertain, unsettled condition of this science of Cetology is in the very vestibule attested by the fact, that in some quarters it still remains a moot point whether a whale is a fish. In his System of Nature, AD 1776, Linnaeus declares, 'I hereby separate the whales from the fish.' But of my own knowledge, sharks and shad, alewives and herring, against Linnaeus's express edict, were still found dividing the possession of the same seas as the Leviathan.
>
> The grounds upon which Linnaeus would fain have banished the whales from the waters, he states as follows: 'On account of their warm bilocular heart, their lungs, their moveable eyelids, their hollow ears, penem intrantem feminam mammis lactantem,'[4] and finally, 'ex lege naturae jure meritoque.'[5] I submitted all this to my friends Simeon Macey and Charlie Coffin, of Nantucket, both messmates of mine in a certain voyage, and they united in the opinion that the reasons set forth were altogether insufficient. Charlie profanely hinted that they were humbug.
>
> Be it known that, waiving all argument, I take the good old fashioned ground that the whale is a fish, and call upon holy Jonah to back me.'

To Melville's Ishmael, a whale is *in fact* a fish: Linnaeus's definition of fish versus mammals does not tally with reality as he sees it, just as most people, even today, choose to believe that oaks are more similar to pines than they are to cacti, and to categorise them as such, even though this runs contrary to the taxonomies of biologists (Dupré, 1981; see below for further discussion). The idea of ecological privilege is easy to state; it is far less easy to justify.

Historically, when it comes to looking at concepts *ecologically*, the distinction between categorisation and analogy at the heart of this investigation is a recent one (just as the

---

[4] "Penetrative reproduction, and females who provide milk by means of breasts" - Melville clearly adhered to the Victorian dictum that delicate matters were best discussed in a classical tongue.

[5] "Justly and properly, according to the law of nature" (Thanks to Robin James for these translations.)

*fact* that a whale is a mammal is a recent development in human classification; historically, whales have been fish for a far longer time than they have been mammals!). As Thomas (1984) argues in his detailed account of changes in natural kind categories in England in the period 1500 - 1800, the conceptual revolution begun by Linnaeus resulted in a common biological system based on *heredity*. However, the system that it replaced was based far more on *analogy*. For much of the early modern period, 'the universal belief in analogy' resulted in much of the natural world being categorised and understood by analogy with human social structures, which led to a radically different partitioning of the world from the one we now employ (as instanced by Ishmael's fictional insistence that a Whale is a Fish).

Thomas notes that Bees had Princes, Potentates, Kingdoms and Dominions (Warder, 1716; Rusden, 1679, quoted in Thomas, 1984 p62); they were ruled over by 'a fair and stately bee, having a majestic gait and aspect' (Levett, 1634, quoted in Thomas, 1984, p62). Cranes followed a captain; Rooks had a parliament; Storks and Ants and Beavers were avowed republicans. As Thomas shows, this picture of the natural world fed back recursively into concepts of human society: King Henry VII once ordered the execution of all mastiffs, after they had baited a lion, 'being deeply displeased ... that an ill-favoured rascal cur should with such violent villainy assault the valiant lion, king of all beasts' (Caius, 1576, quoted in Thomas, 1984, p 60)). Further, as Thomas shows, there is little evidence for a view that Linnaean and post-Linnaean taxonomies necessarily *refine* our view of the world: in many cases, these taxonomies replaced subtle categorisations distinguishing on a range of features with a single crude catch-all based on heredity.

The important issue here is not whether the Linnaean way of construing the world is right, or whether other 'pre- Linnaean' conceptual schemes are wrong. If certain conceptual schemes are ecologically privileged, then one would expect historical data to confirm this. But there is little historical evidence that backs any claims for a privileged partitioning of the world, and there is much that seems to directly contradict it. All that different 'ecological' conceptual schemes such as these reflect is the differing attitudes to pre-theoretical ideas of categorisation that they embody (Lakoff, 1987a, provides numerous examples of the way these analogical taxonomies have remined in the language, illustrating how the Linnaean revolution may be less complete than we generally believe). My claim here is simply that if we wish to explain the cognitive processes that actually underpin categorisation, then an important aspect of our explanation will involve describing the cognitive processes that underlie *all* conceptual systems. An account of categorisation should not be constrained by a particular set of pre-theoretical or metaphysical attitudes to a particular set of categories; rather, it should provide a framework that can account for the cognitive processes that determined prior classification systems, and the mechanisms by which our current view of the world evolved.

Ecological constraints cannot provide any principled answer to the circularity inherent in prototype accounts of category boundaries; even as Neisser (1987) argues for ecological constraints in this respect, he notes the limits of their utility: since very few adult concepts rely exclusively on perceptual similarities - the exceptions are few, notoriously colour (Berlin and Kay, 1969; though see Davies, Corbett, Laws, McGurk, Moss and Smith, 1991, for evidence that even colour may be problemattical in this respect) - any account of categorisation that hinges exclusively on perceptual similarities will ultimately prove to be inadequate.

#### 4.4.1.2 Definition And Representation Of Prototype Essences

There are problems inherent in defining and representing prototype essences (reviewed at length by Komatsu, 1992). The following are particularly noteworthy in the light of Wittgenstein's discussion. Firstly, because prototype theories posit representations of central tendencies that are context-free, they cannot explain how levels of family resemblance - relevance weights - are affected by context (an effect demonstrated by Barsalou, 1985; Roth and Shoben, 1983). Secondly, not all concepts have prototypic characteristics: Hampton (1981) showed that whilst certain abstract concepts (e.g. 'science', 'instinct', 'work of art', 'belief') do have prototypic characteristics, many others do not. Thirdly, prototype accounts cannot account for the kind of knowledge people have about concepts: people know about relations between attributes, and not just attributes alone, which allows them to make reasonable guesses about the meaning of new terms on the basis of exposure to a single instance (Holland, Holyoak, Nisbett and Thagard, 1989; Malt and Smith, 1984). Finally, whilst prototype accounts assume that a concept only represents information about the central tendencies and weight of attributes for a given category, there is evidence that participants who learn a category also learn information regarding the variability of instances of that category (Barresi, Robbins and Shain, 1975; Homa and Vosburgh, 1976).[6]

#### *4.4.1.3 Wittgenstein And Prototypes*

Wittgenstein described family resemblances as playing with words. Moreover, as I noted in chapter 3, taken together, his arguments amount to a rejection of categories as a unitary theoretical construct. There is nothing in prototype theory that, at present, or in principle, indicates that prototypes can or could serve as the basis for a unitary category representation: prototypes posit category boundaries, but can neither justify where they should be, nor why they should be there; they posit that people represent only information about the central tendencies and weights of attributes for a given category, yet cannot account for the knowledge people have about concepts, nor the categorisation inferences people can make; and, as Gleitman, Armstrong and Gleitman (1983) demonstrate, the insight at the heart of prototype theories may not even be any indicator of category structure or representation.

Wittgenstein did not claim that family resemblances are what makes *game* a category; I have tried to show that in the *Philosophical Investigations* he attempted to demonstrate how empty this view is as a definition. Wittgenstein asks, if family resemblances are the common thing that run through game, just as overlapping fibres are the common thing that run through a thread, then what is this thing supposed to be, and how is it supposed to do whatever it is it is supposed to do; *How long is a piece of string?* Prototype theories do little more than attempt a passing answer to the first question - specifying central tendencies and weights of attributes for categories amounts to no more than measuring any number of lengths of string - they provide no insights into the deeper questions posed by Wittgenstein. Nor, in the light of his challenge, do prototype theories provide any compelling reasons to maintain the pre-theoretical attachment to concepts and categories that are determined by unitary representations, in the way prototype theories propose.

---

[6]Another argument against prototype theories claims that there is evidence that prototypes are not indicators of conceptual structure. Gleitman, Armstrong and Gleitman (1983) show that it is possible to get typicality effects even for "clearly defined" concepts such as *odd number*. Moreover, these judgements will predict reaction times in the tasks usually considered indicative of prototypicality. However, it is debatable whether this evidence should be considered as indicative of a problem with prototype theories, or whether it reveals that even "clearly defined" concepts such as *odd number*, or *triangle* have a more complex representational structure than one might assume; and that one's assumptions about their being represented by simple, clear definitions might be far from accurate.

## 4.5 Theories As Essences

Another approach which aims to provide an account of categories as determinable constructs is the 'theory' or explanation based view of categories (Carey, 1985; Medin and Murphy, 1985; Gelman and Markman, 1986; Keil, 1989; Medin and Ortony, 1989; Gopnik and Meltzoff, 1997; see also Komatsu, 1992 for a review). Medin and Ortony's influential treatment of the 'theory' theory is particularly relevant to this discussion, since it claims to argue for a theoretical treatment of categories in direct response to Wittgenstein's arguments.

Medin and Ortony - adopting the traditional psychological interpretation of Wittgenstein - feel that Wittgenstein's "don't think, but look" advice may have been taken too literally. Criticising prototype theory, and the problems inherent in equating category concepts with undifferentiated clusters of properties noted above, Medin and Ortony argue that an adherence to "don't think, but look" involves the abandoning of what is, they claim, the important notion that category membership might be dependant upon intrinsically important (even if relatively inaccessible) features.

Medin and Ortony argue that this may, theoretically, be far too much to bear,[7] and instead they advocate a re-evaluation of Wittgenstein's claims, arguing that:

> on the basis of readily accessible properties that can be *seen*, people presumably will not judge whales to be similar to other mammals. However, if they *think* about the **fact** [*my emphasis*] that whales are mammals not fish, they will probably acknowledge that with respect to some important, although less accessible property or properties whales are similar to other mammals. If one cannot appeal to hidden properties, it is difficult to explain the fact that people might recognise such similarities... there might be a price to pay for looking rather than thinking. (1989, pp 179 - 180)

Medin and Ortony's claim is that Wittgenstein's arguments do not give enough weight to the consideration that similarity judgements are made between representations of objects, and not between the actual objects themselves, since presented objects are perceived and interpreted in terms of concepts that are already known to the individual making the similarity judgement.

[7] That is, too great to be borne by a theory of category representation.

Accordingly, Medin and Ortony argue in favour of a treatment of representations which is rich enough to allow them to show how similarity performs its various useful functions in categorisation. Taking the similarity between whales and bears (being mammalian) as an example, they argue that only an account of representations which embraces the notion of both deep and surface similarities

> can account for why people might believe that two things with very different surface properties (e.g. whales and bears) still are instances of the same category, and therefore why they might judge them more similar to one another than they would on the basis of surface properties alone. (1989, p 180)

This notion of deep similarity is central to Medin and Ortony 's thesis of *Psychological Essentialism*, which holds that similarities between surface features (the kinds of properties listed by participants in typicality experiments, e.g. Rosch & Mervis, 1975) are often constrained by deeper core elements of concepts, which might be termed essences. According to Medin and Ortony, these deeper constraints are vital to understanding how concepts cohere, and hence if one specifies concepts purely by reference to surface features, one is missing out a vital element of conceptual representation.

Medin and Ortony argue that the objections which have turned philosophical essentialism into a position of embarrassment need not apply to psychological essentialism. Traditionally, philosophical essentialism has been dogged by the objection, first noted by Aristotle, that what makes something what it is cannot be independent of how that thing is described. If one wishes to argue that an object x has an essence E, in virtue of which it is an X, and not a Y, then one faces the problem of x's Xness being contingent upon its description D. The same object might be a rock, a paperweight, an ashtray, a door-stop or a weapon depending upon how it is described, thus x will need to have as many essences as there are true descriptions D (or to put in another way, E is essential to D rather than x). All of which severely undermines the notion of there being a particular E in virtue of which x is an X.

Psychological essentialism attempts to avoid this pitfall by denying that the essences it posits have any metaphysical status. Rather than maintaining that objects have essences, Medin and Ortony argue that people believe that they do, and that their representations of things reflect this belief;

> since a major task for cognitive psychology is to characterise knowledge representations, psychological theories about them have to be theories of psychological reality, not metaphysical reality. Thus, if people believe that things have essences, we had better not ignore this fact in our theories of knowledge representation. (1989, p 183)

There is some evidence that people might ordinarily believe that things have essences: Rips (1989) showed that participants are unwilling to change classifications when surface qualities are changed, because they believe that category membership depends upon further intrinsic properties; our ordinary scientific culture embraces a form of essentialism with its claims to be getting at the 'underlying reality' of phenomena. Accordingly, Medin and Ortony do not claim that objects have essences, or that people believe that they know what these essences are, but rather they insist that in spite of this people *believe* that objects have essences, and that cognition is affected by this belief.

Medin and Ortony maintain, however, that psychological essentialism is not a return to classical necessary and sufficient condition categorisation by another means. According to psychological essentialism, people believe that concepts have necessary and sufficient conditions because they believe that concepts have an essential nature, but they do not necessarily believe that these necessary and sufficient conditions *are* a concept's essential nature. "Furthermore, the essential nature may not generate necessary and sufficient conditions at all" (Medin & Ortony, 1989, p184); i.e. it might be believed that being able to fly is part of the essence of a bird, even if not all birds fly.

### 4.5.1 A Knowledge Representation Scheme For Psychological Essences

#### 4.5.1.1 The Essence Placeholder

Medin and Ortony propose that a key element in people's representations of concept knowledge is what they call an *essence placeholder*, which might contain a selection from:

- necessary and sufficient conditions for concept application;
- a more complex, even if potentially incoherent "theory" relating to the concept;

- a belief that experts may know, or be trying to establish, what the essence of a given concept is.

The beliefs contained within the essence placeholder act to constrain the use of surface properties in making similarity judgements.

#### 4.5.1.2 The Identification Procedure

A common defence of the classical account of categorisation has been to argue that categorical essentialism embraces not only core essences, but also an associated identification procedure, i.e. a set of protocols used for relating accessible surface properties to perhaps less accessible underlying essences. Thus, goes the line of this defence, typicality criticisms of classical categorisation apply to these identification procedures, not to the underlying essences (though given that it is identification procedures which govern actual classification decisions, this defence can never explain why essences shouldn't fall foul of Occam's Razor, given their negligible explicatory impact in such accounts of categorisation).

Although this approach falls a long way short of offering a defence for classical categorisation, Medin and Ortony believe that a similar line of argument can help in explaining and strengthening the psychological essentialist position. They argue that if we abandon the messy metaphysical realism of classical categories in favour of psychological realism - the idea that linguistic practices tend to embody a naive metaphysical realism even though this may not be sustainable in fact - then identification procedures can play an explicatory role in psychological essentialism. This goes as follows: embodied in our linguistic assumptions about the world is the idea that, for example, there are connections between being male and typical male characteristics (facial hair, height and weight differences as compared to women), and that these connections result from deeper underlying principles, even if we do not have access to those principles. Later, discoveries by geneticists confirm the realist suspicion embodied in our linguistic and categorical practice.

Thus in Medin and Ortony's account, identification procedures are not distinct from the underlying category essence, but rather they form a continuum, with surface identification procedures and realist beliefs at one end, and, say, genetic theories about the deeper underlying principles for maleness which bear out these beliefs at the other (and presumably, although Medin and Ortony do not make this explicit, in some cases, underlying theories/principles might also act as an identification procedure).

Identification procedures are used in everyday categorisation, however, it is with reference to the deep underlying principles that categories are made coherent.

### 4.5.2 Essentialism And Realism

A deep problem to note in Medin and Ortony's argument is that despite their claims to the contrary, in practice, in order for psychological essentialism to validate *facts* about categorisation in the way they want it to do, Medin and Ortony over-stress the relationship between psychological and metaphysical realism. It can appear, from their account, that there is a strong link between psychological essentialism and philosophical essentialism. In stating:

> our third tenet is that organisms have evolved in such a way that their perceptual (and conceptual) systems are sensitive to just those kinds of similarity that lead them to deeper and more central properties. Thus whales, as mammals that look more like fish than other mammals, are the exception that proves the rule: Appearances are not usually deceiving... psychological similarity is tuned to those superficial properties that are likely to be causally linked at a deeper level. This is particularly likely to be true of natural kinds. (Medin and Ortony, 1989, p 186)

Medin and Ortony seem to be seeking to ground psychological realism in a wider metaphysical realism. Unfortunately, as I argued above, the whale is not an exception that proves any rule. Whales are mammals because of the conventions of biological taxonomy, which picks on certain causal factors (biological heredity) in its classifications.

However, biological taxa fit poorly with our everyday natural kinds (Dupré, 1981); and it is unlikely that Medin and Ortony's account (which tilts in the direction of Putnam and Kripke's naturalistic theories of semantics (Putnam, 1975a; 1975b; Kripke, 1972)) can account for such massive disparity. For example, the various things that we normally classify together as 'trees' belong to a number of different species; pine trees are conifers, whereas *in fact* oaks are angiosperms (flowering plants). Medin and Ortony's story might still hold water if angiosperms and conifers were *in fact* both subspecies of tree, but they are not. Daisies and roses, the former being by no stretch of the imagination trees, are also angiosperms.

To return to the fictional example of Melville's Ishmael, raised earlier: to Ishmael a whale is *in fact* a fish; the biological definition of fish versus mammals does not tally with reality as he sees it, just as to most people, naive psychological reality causes them to believe that oaks are more similar to pines than they are to cacti. Medin and Ortony's position is that 'psychological similarity is tuned to those superficial properties that are likely to be causally linked at a deeper level'. The problem lies in deciding what counts as an *essential* causal link. To Ishmael, the key causal links are those which bind 'sharks and shad, alewives and herring,' to whales, and not those that connect whales to bears. Ishmael may or may not have a theory about fish, however, it is clear that he does not have the further belief (with which Medin and Ortony seem to wish to endow him in order to accommodate a philosophical position put forward by Putnam, 1975a; 1975b; and Kripke 1972) that experts may know, or be trying to establish, what the essences of his concepts are. Where he is aware of an expert, in this case Linnaeus, he seems in no mood to cede the determination of his concept to him. Like the participants reported in Rips (1989) his underlying theory is resistant to transformations of certain superficial properties in the objects in his classification; except that in this case, those properties he considers superficial are those biologists consider essential.

Clearly there are common causal reasons for the similar structures of pines and oaks. It is just that these are not captured by our biological conceptions of reality. Biological conceptions of reality (a descendent of the Linnaean view discussed above, which conforms loosely to what Nagel, 1986, describes as the *physical* conception of reality) consider the causal factors which link oaks and pines to be less essential than the causal factors which link oaks and cacti. Any object might have a number of causal essences, since differing causal factors may have resulted in that object possessing numerous different features. To a biologist, features such as flowering and heredity make roses and cacti and oaks share a causal essence. To others, the environmental stimuli which resulted in the similar surface properties in oaks and pines - sap, woody trunks, etc. - are more essential.

In this light, it becomes difficult to see what distinguishes causal essentialism from philosophical essentialism (described by Medin and Ortony as 'a philosophical orphan, banished to the netherworld of Platonic forms'). Strongly reminiscent of philosophical essentialism, psychological essentialism cast in these terms - i.e. without making concomitant claims to philosophical essentialism - will result in a situation where an object will have as many causal essences as there are different true descriptions of it. Thus psychological essences share with philosophical essences the embarrassing property of being non-essential.

This theoretical conclusion is further backed by evidence from results from two of the few studies to subject the 'theory' theory to empirical scrutiny, Braisby, Franks and Hampton (1996) and Malt (1994). Braisby, Franks and Hampton showed that participants employed natural kind terms in a way that is not consistent with essentialist expectations and predictions; natural kind terms were used in ways that were sensitive to context, and which revealed patterns of apparent self-contradiction. Braisby, Franks and Hampton presented participants with counterfactual scenarios in which the various 'essential' properties of *cat, water, tiger, gold, bronze, lemon* and *oak* were 'scientifically revealed' to be something other than participants might have previously imagined, and then measured the effect that this had on participant's category judgements. Braisby et al conclude that their participants responses reveal that:

> 'the conventional content and use of natural kind terms varies systematically with context ... and not that conventional content and use are invariably associated with essences, as predicted by essentialism.'

Malt (1994) too provides empirical evidence that casts doubt on the central idea of essences as bases for category representation. Malt's study found that the assumption (Putnam, 1975a) that $H_2O$ is the essence of water did not stand up to empirical scrutiny, and that judgements of the amount of $H_2O$ in a liquid were very poor predictors of whether it was water or not.

This does not mean that one has to reject entirely the intuition that the way that we see the world is somehow connected to the way that the world is. Boyd (1989; 1990) argues that causally important properties are 'contingently clustered', not according to necessary and sufficient conditions, but probabilistically. The clustering of such terms, on Boyd's account, results not from mere coincidence, but rather from a causal 'homeostasis', where the presence of some properties tends to favour the presence of others, or where there are properties which underlie certain clusters. These clusters can be loose or tight, and they pick out natural kinds to the extent that they are tight. The resultant 'kinds' will be natural to the extent that they are causally important in explaining phenomena associated with kinds picked out by the cluster. Thus there might be multiple causal clusters underlying a number of roughly similar things: one set of causal clusters underpins properties like sap, woody trunks, branching patterns etc. of trees, in contrast to the cluster of causal features which determine definitions of angiosperms. According to an account like this, the *fact* that *a whale is a mammal* is no more reason to consider whales more similar to bears than fish than the *fact* that *a whale is a fish* is a good reason to think whales and 'sharks and shad, alewives and herring,' more similar than whales and bears: both facts are contingent upon which homeostatic cluster is considered essential.

### 4.5.3 Non-Realist 'Essentialism'?

Medin and Ortony do not need to embrace metaphysical realism. Their original criticism of Wittgenstein (or at least the traditional interpretation of Wittgenstein) was that the comments they attribute to him do not give enough weight to the consideration that similarity judgements are made between representations of objects, and not between the actual objects themselves, since presented objects are perceived and interpreted in terms of concepts.[8] The ideal behind psychological realism is non-realist at heart, since it considers that how people believe the world to be is more important than how the world might really be. Thus what is important is people's representations of things, not the things themselves. Such an account does not need to take any problematic realist notions on board. Causal 'essences' can be only occasionally, contingently, essential, as in Boyd's model, or else they are merely old fashioned essentialism by another name. Indeed, it is hard to see how they could be otherwise: consider Ishmael's categorising whales as fish, which clearly contradicts the 'natural' fish and mammal categories. This is just one example of the many categorisation judgements that individuals make which run counter to the realist 'facts' as established by science. It is not easy to see what role metaphysical realism is supposed to play in *psychological* explanations of the judgements made by these individuals in these case. Psychological explanations of categorisation should be concerned with explaining the psychological processes involved in given categorisation decisions - for example, providing a psychological account of why a given individual might categorise whales as fish - rather than attempting to address the metaphysical validity of such decisions.

---

8 Medin and Ortony (1989) cite no evidence in support of this claim; and I can find nothing in the text of the *Philosophical Investigations* that suggests that Wittgenstein viewed conceptual judgements as determined by anything other than the interactions of mental representations. However, it is worth noting that Wittgenstein had strong views on what *form* such representations *could not* have.

Psychological essentialism is based upon the intuition that people believe that objects have essences, and that this belief about essences affects our interactions with the world. It is not based upon the idea that these beliefs (however true the facts of the matter of their being held may be) are *true beliefs*. It can be true that people have the belief without the belief being true, and the causal effects of the former are independent of the latter. Medin and Ortony accept in their knowledge representation for psychological essentialism that theories (as contained within an essence placeholder) may not be coherent. Rather than adopting a realist standpoint, and arguing that evolution must have prepared us to see features indicative of objects' essences, one might be better off adopting a more non-realist (yet psychologically realistic?) perspective. The realist position - of whatever persuasion - would lead one to expect to find coherency in people's theories about essences; the evidence points to inconsistencies and incoherencies in everyday notions of essences. A realistic, non-realist approach would address the question of how it is evolution has provided us with the wherewithal to recognise and pragmatically make use of causal clusters, which may contradict amongst one another, according to the contexts in which we find ourselves.[9]

If we return to Medin and Ortony's knowledge representation, the belief, contained within an essence placeholder, that experts may know, or be trying to establish, what the essence of a given concept is, is at best superfluous (witness Ishmael's - albeit fictional - scant regard for the labours of Linnaeus), and at worst reintroduces realism through the back door. Recasting psychological essentialism in the non-realist terms which seem most appropriate to it, an essence placeholder would contain a selection from:

- conditions for concept application;
- a causal, even if potentially incoherent "theory" relating to the concept.

The further belief posited by psychological essentialism:

- that experts may know, or be trying to establish, the essence of a given concept

---

[9] Such an approach would also - essentially - treat the question of conceptual (or category) 'essences' as a matter for empirical investigation, rather than assuming them at the outset as a theoretical imperative.

would be rejected as superfluous - individuals may have such beliefs, but it is hard to see how these beliefs can give their mental representations of concepts a particular coherent structure.

### 4.5.4 Theories And Essences - What Is Essential?

As I detailed above, however, Wittgenstein's arguments in the *Philosophical Investigations* amount to more than a rejection of necessary and sufficient conditions as a basis for categorisation (the traditional psychological view): in the course of his discussion, Wittgenstein presents a number of substantial objections to essences and schemas (whether "theoretical" or not):

- Firstly, the lack of necessary and sufficient conditions for category membership makes essence or theory definition problematical, to say the least (the point Medin and Ortony acknowledge and seek to address);
- Secondly, there is the question of what form an essence representation or schema (even one that is only psychologically essential) should take: *what* theory are people supposed to have for games?;
- Thirdly, there is the question of what is essential to a psychological essence? Which aspects of a representation of a psychological essence are generic, and which are not, and how do we explain the distinction between generic and non-generic aspects of the representation in the use of the proposed essence?
- Finally, what is supposed to be intrinsic to the representation of a psychological essence that would cause it to be used differently to examples of whatever it was that it was supposed to be definitive of?

Whilst Medin and Ortony address the first of these points - identification procedures don't specify completely, theories can be partial, or even incoherent - nothing in their account - or in the more general literature advocating explanation of theory based accounts of categories - addresses how theories, explanation schemas or psychological essences are supposed to overcome Wittgenstein's other three challenges.[10]

[10]There is no doubt that theories *would* provide a useful means of categorising the world. The question is, however, whether theories could actually provide a plausible and coherent account of the cognitive categorisation capacities that people actually employ. As Hahn and Chater (1997a) observe, whilst the rule-based classifications used in science provide spectacular evidence of the predictive and explanatory utility of theories, in fact, very few aspects of scientific knowledge have been adequately formalized. Hahn and Chater note also that constructing theories for common-sense knowledge appears to be an

In this light, the question at hand becomes one of finding a framework that can accommodate the important insight at the heart of psychological essentialism: the idea that relations between attributes in representations are important; especially causal relations in representations. Prototype theories refer only to undifferentiated, probabilistic clusters of features, and are notoriously bad at accounting for the knowledge people have about concepts, and the categorisation inferences people can make. Including causal relations in people's representations of conceptual knowledge could begin to account for these factors. The easiest way to account for the fact that causal beliefs may (in some way still to be determined) have some causal repercussions cognitively, without enmiring itself in the dubious notion that realist beliefs may be true in reality, would be to reject the idea of concepts as unitary, determinable constructs, or even, ultimately as essences. It is hard to see how the contents of the modified essence placeholder could be sufficient to account for categorisation without reference to example information as well (if only to deal with exceptions, see Barresi, Robbins and Shain, 1975; Homa and Vosburgh, 1976). If we accept that representations of examples of objects will contain important relational information, then we can accommodate Medin and Ortony's positive arguments without incurring the theoretical - even metaphysical - costs that the 'theory' theory framework brings with it. After all, Medin and Ortony argue that theories may be vague, or even incoherent: in the light of the foregoing, it would appear that insofar as it is possible to defend any positive account of theory theories, theories won't even need to be essential.[11] A defensible 'theory' theory will be of the following, radically weakened form:

---

even more difficult task. Common sense rules seem to almost always have exceptions (Reiter, 1980); even when rules are available, their application to specific instances appears to depend on vast amounts of background knowledge (still further theories?) in ways that are still poorly understood (Oaksford and Chater, 1991; 1993). As I noted above, common-sense knowledge does not divide easily into domains, leading to a situation where providing rules for parts of knowledge seems to lead to the seemingly endless task of capturing the whole of human knowledge; reflective of what Fodor (1983) has called calls the *isotropy* of human knowledge (see Hahn and Chater, 1997). It seems that trying to provide a theory for even a simple domain of common sense knowledge could soon escalate into a requirement to provide an account of *all* such knowledge.

[11] It is worth making the observation here that findings such as those by Rips (1989), showing that participants are unwilling to change classifications when surface qualities are changed, because they believe that category membership depended upon further intrinsic properties, may tell us more about what participants *believe* is essential to concepts than they do about the *processes* that actually determine the concepts participants employ.

1. Theories are representations of clusters of causal information;

2. These theories will be partial.[12]

Despite the claims of theory theorists, these two points - which I have argued are the only elements of the 'theory' theory that can be given a principled defence - appear to amount to little more than a claim that causal information is important to the representation of categories (this is a point I shall return to at length later in this thesis).

Just as Wittgenstein did not claim that family resemblances are what makes *game* a category, so he also attempted to demonstrate that neither did any theory of game make *game* a category. Most of the concepts used in everyday cognition do not have the coherence a strong theoretical claim would imply. The claim that 'partial theories' or background knowledge - i.e. the inclusion of causal or other relations in conceptual representations - are relevant to categorisation need not conflict with Wittgenstein's arguments; insofar as it is possible to ascertain how these would be worked out, they do not contradict Wittgenstein's theoretical account of the role examples play in the acquisition and use of concepts and categories. Firstly, as I have sought to show above, insofar as 'theory' theories can be defensibly stated, they make no statement about boundedness, nor do they make any claim regarding defining features. Secondly, though the theory-based view does suggest that learning and understanding a category also involves acquiring appropriate causal knowledge, again, this does not directly contradict Wittgenstein's account of examples in acquisition and use but merely suggests some additional factors that should be considered in addressing the representation of examples of concepts, and the processes that govern the usage of these examples.

As with prototype theories, when analysed, theory theories provide few, if any compelling reasons to maintain a pre-theoretical attachment to concepts and categories that are determined by a unitary representation, or indeed by any unitary - even 'theory-like' - theoretical construct.

---

[12] Given the problems inherent in definitional accounts of conceptual structure - see above - one must assume that 'theories' are not complete; i.e. they inform classification decisions, but are only 'partial', in that they form one component of a complex, non-deductive overall process (Hahn & Chater, 1997). Ramscar and Hahn (1998b) note that this does however, pose a potential problem regarding partial theories, i.e. how partial does a theory have to be to not be stating an 'essence'? This is a moot point, but given that the theory-based view has done little to provide full accounts of any categories, whilst no definitive answer can be given to this question, the onus of providing an account and spelling out its implications seems to be firmly with the 'theory' theorists.

## 4.6 Examples Of Things

The evidence and arguments above lend some credence to the third part of the Wittgensteinian account of category representation that I established in chapter 3 - that learning a category such as game does not involve learning or extracting any essence or schema from the instances of the category one encounters. Logically, this would also, in itself, appear to lend support for the fourth part of the account, since without a central schema to act as a category representation it is hard to see how else one could learn a "category" such as game, except by learning examples of *game*, and appropriate ways of using these examples.

Arguments of the 'how else could it be?' kind are, however, rarely compelling or convincing. Having shown that there is empirical evidence against 'central schemas', what is needed is positive evidence in favour of an exemplar account. A point related to this that has eluded many discussions of conceptual storage is that debates between single prototypes versus exemplars revolve - to an extent - around questions of granularity rather than principle. As Borges (1962) succinctly demonstrates in the much quoted *Funes the Memorius*, even an 'instance' is a generalisation of sorts. When Funes struggles to

> comprehend that the generic symbol *dog* embraces so many unlike individuals... it bothered him that the dog seen at 3.14 (seen from the side) should have the same name as the dog at 3.15 (seen from the front).
> Borges (1962) pp 93-94

he encounters the following problem: if the concept of *dog* determines whether things we encounter are dogs, then in the same fashion the concept *Spot* must similarly unite a certain class of experiences of a particular dog as being experiences of Spot the dog, and so on (a similar point is made by Barsalou, Huttenlocher and Lamberts, 1998): they argue that it is important to note that categorisation must take place on the basis of both individuals and events:

> "If the cognitive system didn't establish representations of individuals that exist across events, it couldn't construct the history of an individual, it couldn't represent the fact that the appearance of an individual might vary widely across occasions, and it couldn't count the number of repeating individuals observed across occasions, and it couldn't determine the

properties that occur most often across the individuals in a category. Establishing representations of individuals captures the physical structure of the world, such that important inferences about the entities in it are possible.

In contrast, if the cognitive system recorded only information about events, it couldn't distinguish individuals that occur frequently in a category from individuals that occur rarely. Similarly, it couldn't distinguish the frequent properties of an individual from the infrequent ones. In general, the representation of events captures what is likely to happen to an agent in his or her experience. Whereas frames for individuals capture what exists in the world, event memories capture how the world is likely to affect an agent in a given event"

Barsalou, Huttenlocher and Lamberts (1998, pp 257).

Once factors such as these are considered, it seems reasonable to argue that one of the key questions of conceptual storage is best viewed not as being a question of instances versus generalisations, but rather as being one of deciding between *unitary* versus *multiple* representation accounts of conceptual storage. Unitary accounts of categorisation posit a single stored representation - schema, prototype or, perhaps, theory - in virtue of which items are classified into a category as the outcome of some process. Multiple representational accounts posit the storage of a number of representations (perhaps at different levels of granularity, from 'instances' to broad intermediate generalisations) which may jointly or individually result in some object being categorised - whether as an individual or an event - as the outcome of some process ; one way to interpret the observations of Barsalou, Huttenlocher and Lamberts (1998) is that all categorisations will involve a process that compares representations of *both* individuals and events..[13]

These different models of conceptual representation will have different implications for theories of categorisation. If a unitary representation model were correct one would expect that provided one could specify the stored representation and the process by which objects were related to it, one should in principle be able to give a definitive

[13]There is, however, no need for such processes to regress terminally: as Malt (1995) argues, some in-built regularities - perhaps as a result of perceptual constraints - will ultimately be inevitable.

account of, say, why it is some things are X's. I have attempted above to show how unlikely it is that any account such as this will be forthcoming.

On the other hand, a multi-representational account will most likely not admit any definitive account of X's at all, since the multiplicity of relations between the differing stored representations that could influence an object's X-ness will most likely, insofar as it is possible to predict, preclude any kind of general account. A specific object's X-ness might be dependant upon that - and only that - objects' interaction with a particular subset of the stored elements relating to X-ness, and the concomitant process by which X-ness is adjudged.

Clearly this last position is in concord with Wittgenstein's final arguments, which reject the idea of some abstracted schema in preference for an account based on previously encountered examples and appropriate ways of using these examples. Variants of this view - usually referred to as the *exemplar*, or instance-based approach - have been much explored in recent attempts to model the cognitive processes underlying categorisation. Like the prototype approach discussed above, exemplar theories also see categorisation as a process that involves comparisons to stored representations. However, rather than category membership being based upon a judgement of an item's similarity to a single stored representation of each category, as posited by prototype theories, in exemplar theories, objects are instead compared to many stored exemplars (previously encountered instances) of a category (Nosofsky, 1986; 1987; 1988; 1989; 1991; Nosofsky, Clark and Shin, 1989; Shin and Nosofsky, 1992; Lamberts, 1996; Medin and Schaffer, 1978).

Thus where the exemplar, or instance-based, approach differs from prototype theory is that abstraction across instances (i.e. schema abstraction) is not assumed to take place at or soon after the time of 'concept acquisition', independently of concept use, but rather to take place at the time concepts are used. Abstraction across instances is presumed to occur dynamically at the time that membership judgements, typicality judgements or conceptual inferences are made (Hahn and Chater, 1998; Komatsu, 1992).

### 4.6.1 The Generalised Context Model (GCM)

With regards to exemplar-based psychological models of categorisation, Nosofosky's (1986) Generalised Context Model (GCM), has provided remarkable data fits to human behavioural data in a variety of experimental contexts (e.g., Nosofsky, 1986, 1988). In keeping with the general principles of exemplar models described above, this model assumes that categories are represented in terms of stored exemplars.

The workings of the GCM are summarised in Ramscar and Hahn (1998b): exemplars are represented as points in a multi-dimensional "psychological space"; co-ordinates for points are determined by their value along the particular psychological dimensions in question (these dimensions can be things like "loudness" or "size" or more complex, composite dimensions); and similarity between exemplars is a function of distance in psychological space (specifically an exponential decay function, see Nosofsky, 1986).

Classification decisions are then governed by a *probabilistic* response rule. The probability of categorising an item as a member of a particular category corresponds to the weight of the evidence in favour of its membership of this category. In contrast to nearest-neighbour algorithms, the evidence weighing process takes *all* exemplars into account, and all exemplars are treated as equal for the purposes of supplying evidence. Specifically, the strength of the evidence for a category C corresponds to the summed similarity between the novel item and all known exemplars of C, divided by the summed similarities to all stored exemplars; that is, the evidence weighed takes in not just members of C, but also the members of relevant competing categories.

As a consequence, the model does not impose discrete category boundaries in psychological space, but rather *probability distributions* over the entire space.

### 4.6.2 Evidence For Exemplars

I have noted a number of times that most experimental results addressing categorisation do not directly indicate *anything* about conceptual representation: separating form, content and the processes acting on concepts is an invidious business (perhaps best illustrated by Wittgenstein's detailed remarks on the use of schemas, discussed in Chapter 3, above). However, the *issue* of whether or not a particular learning process involves the abstraction of some core essence or not - be it schema, theory or prototype - has been central to much recent experimental psychology and, as Ramscar and Hahn (1998a) observe, has not only been pursued in concept learning tasks, but also in related domains such as Artificial Grammar Learning (Shanks and St John, 1994).

The whole issue of the abstraction of category information from instances has been clouded in controversy. Not only is the empirical evidence for or against abstraction contested, but there has also been much disagreement regarding the conceptual and empirical criteria on which a distinction between exemplar storage and abstraction in category representation can - in principal - be based. Barsalou (1990) has argued that a principled distinction between exemplar storage and abstraction in category representation is impossible, whilst Hahn and Chater (1998) have argued in turn that Barsalou's position is based on a highly idiosyncratic notion of abstraction. Hahn and Chater's evaluation of the many criteria that have been put forth, particularly in order to distinguish between processes based on rules and processes based on exemplar similarity, reveals - in support of Barsalou's claim - that many proposals have overstated their ability to cleanly distinguish between the two. However, whilst acknowledging the problems inherent in this area, Hahn and Chater (1998) argue that from an experimental perspective, the most straightforward - and pragmatic - way to address the issue of abstraction is through model to model comparisons of fully specified cognitive models.

Ramscar and Hahn (1998a) note the following evidence for regarding prototypes as an example of these problems:

> "evidence for prototypes in natural language categories has been sought from a variety of sources. Classic are those studies which identified a variety of so-called "prototype effects"; all of these involve some form of differential reaction to central or typical members of a category such as differences in typicality ratings, faster reaction times in speeded classification tasks or differential retention in memory relative to other items (see e.g. Rosch, Simpson and Miller, 1976; Posner and Keele, 1968; Posner and Keele, 1970). However, such effects do not unequivocally indicate *mental representations of concepts* in terms of prototypes (Lakoff, 1987b). Rather such effects might arise from cognitive representations and processes which make no use of representations of prototypes or central tendencies as such." (Ramscar and Hahn, 1998, pp 397).

These points receive support from those studies which have attempted to compare the performance of fully specified categorisation models. The literature on categorisation contains a number of studies in which model comparisons between exemplar models which simply store all encountered instances in memory, and prototype models which abstract a central tendency have been conducted. These comparisons have consistently favoured exemplar models. Exemplar models have yielded quantitative fits superior to the prototype models tested and accounted for a wide range of phenomena traditionally associated with prototypes such as the instability of instance retrieval and typicality judgements; the levels of specificity at which concepts are encoded; sensitivity to correlations amongst category instances; and the way accuracy in classification tasks increases with category size (Nosofsky, 1986, 1987, 1988b, 1989, 1991b, Nosofsky, Clark and Shin, 1989, Shin and Nosofsky, 1992; Lamberts, 1996).

Similarly, as mentioned above, those few empirical studies that have directly addressed the assumptions behind core essences - whether as schemas or theories - have found little or no support for the idea that essences are extracted in category learning. For example, Malt (1994) found the idea, put forward by Putnam (1975), that $H_2O$ is the essence of water did not stand up to empirical scrutiny, and that judgements of the amount of $H_2O$ in a liquid were very poor predictors of whether it was water or not.

In summary, at present at least, there is no clear evidence in the literature for abstraction in the acquisition of concepts, whilst there is considerable evidence which can be marshalled in support of some kind of exemplar-based account.

### 4.6.3 Kinds Of Similarity

One main criticism of existing exemplar models (Medin, 1989) is that they use a very limited form of similarity metric, considering only features of representations, and not relations between features (Medin, 1989; Medin, Goldstone and Gentner, 1993; Goldstone, 1994). Whilst exemplar models have been shown to out perform prototype models in modelfitting comparisons, these comparisons have relied on artificial test domains and data, whose conceptual richness is questionable, often to the point where the application of 'rich' appears to be a serious misnomer. Stimuli will often vary by only a few characteristics, with typical variations taking place between a few of only a handful of attributes per stimulus item (e.g. Barsalou, Huttenlocher and Lamberts, 1998; Nosofsky, 1986; 1988; 1991; Shin and Nosofsky, 1992; Lamberts, 1996; Medin and Schaffer, 1978).

Noting this, it is worth recalling that in my discussion of the 'theory' theory I described the array of evidence indicating that feature-clusters alone are insufficient for capturing and modelling human concepts and categories. This evidence indicates that relations between features will also have an important role to play in the structure and coherence of concepts (insofar, that is, as concepts *are* structured and coherent). At present exemplar models are almost exclusively feature driven, and do not capture the role of relations in their exemplar representations (criticisms about the failure of models to capture relational information could, however, be levelled with equal validity, against *all* existing models of categorisation).

Thus, it seems fair to acknowledge at this point that whilst exemplar models outperform central tendency or prototype models on the artificial tasks common in the literature, these models still appear to fall a long way short of capturing the similarity matrices that are used cognitively in ecologically valid domains.

## 4.7 Summary

### 4.7.1 Prototype Theories.

Accounts of category structure in terms of similarity to one or more central "prototypes" are incompatible with Wittgenstein's perspective and the very real theoretical problems he raises (despite the frequent appeals to Wittgenstein from proponents of prototype theories, e.g., Taylor, 1995). This holds both for versions of prototype theory which view the prototype as an abstracted central tendency or schema, and for those versions which take prototypes to be particular, privileged exemplars (for discussion of the different variants of "prototype" see Barsalou, 1987).

The schema version is irreconcilable with Wittgenstein's position. (Such accounts, seem particularly popular within Cognitive Linguistics - see, e.g., Taylor, 1995 - but have also been proposed in psychology; see for instance., Smith and Medin, 1981, on the "probabilistic view".)

In contrast, the conflict between his position and the "prototype-as-privileged-exemplars" stems from the fact that such prototypical exemplars would, in fact, provide the glue to hold the category together in a way that Wittgenstein denies. Central exemplars *would* constitute a central thread or focal point around which the category is organised. Items would all obtain category membership by virtue of the single, simple fact that they are sufficiently similar to a central exemplar.

This is not the "criss-crossing" associated with Wittgenstein's idea of family resemblance (and indeed not the way real families, viewed over generations, are structured). Thus the popular equation of "the family resemblance" view of category structure, which claims direct descendence from the *Philosophical Investigations*, and "prototype theory" must be rejected.

### 4.7.2 Theory Based Views.

As I noted above, the lack of computational explicitness of 'theory' theories - or explanation-based views - makes a straightforward comparison between them and Wittgenstein's account difficult. Given the notorious problems with definitional accounts of conceptual structure, one has to assume that "theories" are not complete, i.e. that they are sufficient to allow deduction of classification decisions, but that they are only "partial", in that they form one component of a complex, non-deductive overall process. This overall process, however, which could involve similarity comparisons, has not been spelled out by advocates of the theory-based view. As I argued above, the simple claim then that "partial theories" or background knowledge are relevant to categorisation need not conflict with Wittgenstein's critique. Such claims makes no statements about boundedness, nor do they claim definitional features. Though the 'theory' theory suggests that learning and understanding a category also involves acquiring appropriate background knowledge, this need not contradict the role of examples in acquisition and use. Rather, it merely suggests that additional factors that should be considered in modelling an exemplar driven process.

### 4.7.3 Exemplar Theories

To recap: exemplar theories assume that our mental representations of categories consist simply of stored exemplars, i.e., known members. There is no abstraction of schemas or central patterns. Despite what Ramscar and Hahn (1998b) describe as their "very extensional feel", exemplar accounts do allow classification of novel, previously unencountered objects simply by virtue of their similarity to known exemplars. In the most basic version, a novel item is simply given the classification of the known exemplar to which it is most similar, i.e., classification is based on the single "nearest neighbour" in similarity space. Despite this simplicity, it is an approach to categorisation that proves remarkably successful in machine learning contexts (Cover & Hart, 1967). As a cognitive model, however, it conflicts with Wittgenstein's claim that natural language categories have no boundaries. The simple nearest neighbour approach produces well-defined category boundaries which run along the paths of equal distance between members of competing categories.

However, as a class of account, exemplar models need not posit such boundaries and, in fact, the GCM, the most prominent exemplar model in the psychological literature, does not. As described above, classification decisions in the GCM are governed by a *probabilistic* response rule - the probability of categorising an item as a member of a particular category corresponds to the weight of the evidence for this category - and consequently the model does not impose discrete category boundaries in psychological space, but rather it calculates *probability distributions* over the entire space.

The GCM seems to be compatible with all 4 points extracted from Wittgenstein's argument above: first, categories have neither necessary nor sufficient conditions; second, category spaces are unbounded; third, learning does not involve extracting an essence or schema; fourth, in learning a category such as "game", one learns examples and appropriate ways of using these examples. Thus, accepting the caveats noted above concerning the artificial domains in which the GCM has been tested, there is at least one contemporary cognitive account of categorisation which is broadly compatible with Wittgenstein's description of categories and category structure.

## 4.8 The Concept Of 'Concept'

One thing that emerges from the process of contrasting the substantive content of Wittgenstein's arguments with the leading models and perspectives in current categorisation research is a clear demonstration of the continuing relevance of the issues Wittgenstein raised. Wittgenstein's arguments - and the evidence - bear down on any all-encompassing view of category structure. Together, the two appear to effectively explode the idea of the category as a unitary theoretical instrument: how likely is it that, given that categories aren't defining features, shared essences or some other common thread running through, that there is a fundamental unity in all categories? That clear cut members all have higher within category similarity than between category similarity (as predicted by GCM) or that all categories are based on partial theories, etc.?

Ecological, natural language categories - naming is our prime categorisation behaviour - are produced by collectives rather than individuals. Natural language categories develop - evolve - over time and this process of evolution can be subject to diachronic accidents. These are factors which might reasonably be expected to put a bound on whatever systematicity, at whatever level, one might hope to find in concepts and categories. If category structures are like this, it seems reasonable to assume that the naming behaviour of individuals will, to some extent follow, denying cohesion even at the level of processing. If category structures are variable, then it seems to follow that processing might also involve multiple, even competing influences. Cognitive science has made great steps forward by developing and testing constrained models. But if one pauses to look more at the untidiness of our real world categories, Wittgenstein's scepticism about category *structure* does not seem to find any happy resolution in unitary accounts of cognitive *processing*: the deep questions Wittgenstein poses loom large still.

The evidence discussed here does not indicate that ecological concepts and categories are, or can be, neat, bounded entities. Rather, it suggests that concepts and categories are created and evolve dynamically, according to context and task demands (both local and historical), using a mixture of stored representations (examples) and processes (means of applying examples). (And, though I will return to this point later, it is worth remarking here how similar this description sounds to a characterisation of analogy!)

### 4.8.1 A Short Family History

I noted above that natural language - or ecological - categories are collective, rather than individual, products,[14] subject to a range of influences over their evolutionary course. Before considering the way in which the picture of categorisation painted in this chapter fits with the contrast definitions made for analogy in chapter 2, I shall attempt to illustrate some of the points made here by considering the status of a concept that has been at the heart of many of the issues discussed so far: the concept of *family*. What are the family of resemblances in *family*?

Williams (1976) in his classic study *Keywords*, traces the history of the word 'family', and its attendant 'concept'. **Family** was first used in English in the period straddling the end of the 14th century and the beginning of the 15th century. Williams identifies **family**'s forerunner word as the Latin *familia* (which conceptually, was more akin to the modern English *household*), which has its ultimate root in the word *famulus*,[15] which corresponds to the modern English servant.

---

[14] Wittgenstein too had much to say on this point, c.f. McGinn (1985).

[15] Having argued that concepts evolve *diachronically*, I do not wish to imply that the history of **family** pre-*famulus* is *prochronic*; this is just the earliest point to which I can trace this

As Williams notes, the associated adjective **familiar** (which appears to have an earlier common usage) has a range of meanings reminiscent of the more general use of *family* prior to around the middle of the 17th century. Just as in early use of **family**, there is a:

> "direct sense of the Latin *household*, either in the sense of a group of servants or of a group of blood-relations and servants living together in one house. **Familiar** related to this, in phrases like **familiar angel**, or **familiar devil** and the later noun familiar, where there is a sense of being associated with, or serving someone." (Williams, 1976, p108; the emphasis is Williams')

In the fifteenth and sixteenth century, the phrase **familiar enemy** was common, and was used to indicate an enemy within one's household, and by extension, an enemy within one's own people (somewhat akin to the modern idea of a fifth-columnist). However, as Williams notes, the strongest early senses of **familiar** are those that are still with us in modern English: friendship or intimacy; well known; well used; or habitual. These uses were derived from people's experiences of living together in households, in close relations to one another:

> "They do not, and **familiar** still does not, relate to the sense of blood group" (p. 109)

It appears that Family, was then extended, from the end of the fifteenth century, to describe a house; a group of kin. In the Authorised Version of the Bible (1611) **family** was restricted to these wide senses: either a large kin-group, often virtually equivalent to a tribe (*Genesis* 10:15, 12:3; *Jeremiah* 1:15, 31:1; *Ezekiel* 20:32) or the "kin-group of a common father: 'and then shall he (a brother) depart from thee, both he and his children with him, and shall return unto his own family, and unto the possession of his fathers shall he return' (*Leviticus* 25:41; cf. *Numbers* 36:6)." (Williams, 1976, p109)

In none of the senses current before the mid-period of the seventeenth century do we find the distinctive modern concept of **family**: the sense of a small group confined to immediate blood-relations. Indeed, Williams notes that when this concept was needed

---

evolutionary path. (See Gould, 1995, for a lively account of how Philip Gosse attempted to divide evolutionary change as 'prochronic' change - occuring outside of time, and fahioned by God at the time of Creation - and 'diachronic' change, that unfolded in conventional time post-Creation).

in *Genesis* of the Authorised Version of the Bible, *near-kin*, rather than **family** was used.

In the period between the seventeenth and nineteenth centuries, the modern sense of **family**, as a small kin-group, normally living in one house, came to be dominant, to the extent that in the twentieth century, sub-ordinate category distinctions between **nuclear families** and **extended families** have come to be made. This transition has a complex history: Williams notes that in 1631, we can still read "his family were himself and his wife and daughters, two mayds and a man" (Williams, 1976, p109), where **family** is being used in place of the modern *household*. This use survived in rural England "beyond the eighteenth century, and the later distinction between family and servants was, in this instance, very much resented" (Williams, 1976, p110); given the original root of **family** in *servant*, it would appear that one quite literally could be the master of one's family.

Other forms of the concept **family** from around this period include a specific association with children, as in he "duly sent his family and wife" (Alexander Pope, in Wasserman, 1960).

Williams argues that the dominance of the small kin-group idea of family was probably not established prior to the early nineteenth century, where massive social pressures changed the word, and the concepts associated with it. Specifically, Williams argues that the middle nineteenth century was the high-watermark of the **bourgeois family**, and that: "the sense of the isolated family as a working economic unit is clearly stressed in the development of capitalism" (p110). The nineteenth century saw a trend to distinguish between *work* and **family**: a man worked to support his **family**, and his **family** was supported by his work. This form of the concept **family** would have received broader social reinforcement from the development of smaller, individual houses, and as a result, households, especially in the new working and lower-middle classes throughout the nineteenth and twentieth centuries. Whereas in the eighteenth century, to **found a family** - in both the sense of lineage, servants and property - had currency amongst the aristocracy, the use of **family** by the nineteenth century masses had no need for the conceptual attributes of household (such as servants and worries about lineage); instead the near kin-group was sufficient to define the most common social relationships experienced. Says Williams:

> "**Family** or **family and friends** can represent the only immediate positive attachments in a large-scale and complex wage-earning society.

> And it is significant, that class-feeling, the other major response to the new society, used *brother* and *sister* to express class affiliation... it is significant also that this use of *brother* and *sister* came to seem artificial or comic in middle class eyes. **Family**, there, combined the strong sense of immediate and positive blood-group relationships[16] and the strong implicit sense of property" (p. 111).

Thus whilst at first glance it might appear that the use of **family**, **brother** and **sister** by trades unionists is metaphorical, the evidence provided here should give one pause for thought. If the trades unionists' use of family is metaphorical, what is the *literal* meaning of family? The history of the concept family is indicative of numerous acceptable uses, and the extension - and contraction - of the concepts associated with it according to usage and similarity of usage further indicates this.

As Williams remarks at the end of his essay, this family history is worth recalling when politicians talk about the **family** as being an institution that is breaking up; or on the verge of collapse. I argued above that natural language categories are the products of collectives and that they develop over time, over the course of which they are subject to a variety of 'accidents.' The history of **family** appears to provide a fitting example of this. Yes, there *are* family resemblances between the various uses of **family**. But, in keeping with Wittgenstein's analysis, it appears that:

- there are no necessary or sufficient defining characteristics to **family**: rather, kinships - "family resemblances" - can be traced across the concept;
- the category space of **family** is unbounded - i.e. there are no boundaries to the space across which "family resemblances" can be traced (see Williams' remark on the 'collapse of the family' above as an indication of the artificial problems that assumptions about bounding can bring);
- learning the category **family** does not appear to involve extracting an essence or schema from instances (certainly, anything one might put forward as essential - **"familiarity"**? - would be insufficient to capture the particular uses of **family**; and would in turn beg the question of what was the essence of **familiarity**).

[16] This is significant - some working class areas of Glasgow were as much as 80% male in the late nineteenth century (Lister, 1997). In so-speaking, Trades-Union *brothers* were referring to the only **family** they had. Literally.

- learning, and use of the "category" **family**, appears to depend upon examples (instances) and appropriate ways (note the similarities between the various uses of **family** over time) of using these examples.

Moreover, the *variety* of uses of **family** seems to provide support for the 'multiple-' (hence, no-) essence view of concepts that emerged from my discussion of the 'theory' theory above.

Which use of **family** is essential? Which is ecologically privileged?

From the historical data, it would appear that the answer is, none.

## 4.9 Categorisation And Analogy

I have argued that Wittgenstein's arguments and the weight of psychological evidence both bear down on any all-encompassing view of category structure, and that taken together, they effectively undermine the idea of the category as a unitary theoretical instrument. I also argued earlier that in order for the standard contrast definition of analogy to do its work, an account of categorisation, *distinct* from that contrast definition, was necessary. This survey would appear to show that no such account is available, nor, since it doesn't seem likely that any answers to Wittgenstein's deep questions regarding any 'straightforward' account of categorisation will be forthcoming, does it seem likely that such an account of categorisation will be available in the future.

In chapter 2 I described how analogy is defined as being distinct from categories, the nature of which are left unexamined, presumed real and determinable, and argued that without an account of categorisation, characterisations of analogy and metaphor reliant upon a contrast with categorisation will fail to do much characterising at all. Analogy is consistently defined in contrast to categorisation. I argued earlier that a valid contrast definition requires an account of at least one of the contrasting elements. From the foregoing, it would appear that categories can't do the work necessary to provide this distinction.

At the end of Chapter 2, I argued that without an account of categorisation, a definition such as:

> "In an analogy, a familiar domain is used to understand a novel domain in order to highlight important similarities between the domains, or to predict new features of the novel domain." (Clement and Gentner, 1991, p89)

would be more accurately reformulated along the lines of: 'in analogy, a stored representation, or representations are used in order to highlight important similarities between it and a new representation of an object or concept, or to predict new features in the new representation of an object or concept.' This definition seems to be as applicable to an exemplar model of categorisation as it does to analogy.

Theoretically, at least, distinguishing analogy from categorisation may not be the simple task our intuitions - and the literature - might have us believe. Indeed, from the evidence examined above, it would appear that, at present at least, there is no distinction to be made.

## 4.10 Categories Re-Visited - Reaction Time Studies

Objections to the theoretical denial of a distinction between analogy and categorisation I make above hinge largely on the intuition that fundamentally there is something more to categorisation, or literal judgements, as opposed to analogical judgements, namely categories. Whilst it may be hard to characterise what categories are, says this objection, we can determine their existence through their basic effects in cognition, namely through the way that analogical and metaphorical understandings are dependant upon literal 'categorical' understandings, with literal understandings being computed first, and idiomatic interpretations being derived from them (often called two-process theory (Clark and Clark, 1977; Kintsch, 1974)).

Despite the widespread appeal of this view (and its consistency with pre-theoretical intuitions about analogy, metaphor and categories), existing data does not support two-stage theories (Hoffman and Kemper, 1987). Commenting on their comprehensive review of studies which employed chronometric measurements of comprehension processes for non literal language - idioms, indirect requests, proverbs and metaphors[17] - Hoffman and Kemper make the following remark:

[17] Given the foregoing, I shall treat terms such as literal, idiom, indirect request, proverb and metaphor in this section as pre-theoretical constructs. The cognitive and methodological validity of distinctions between 'literal' and 'non-literal' meanings made in these experiments is questionable to say the least;

> [*On indirect requests and idioms*] "Overall, the research on comprehension of indirect requests and idioms shows a remarkable degree of agreement despite differences in materials and methods. Such nonliteral forms are comprehended more rapidly than if each word were literally processed. The experiments suggest that people do not analyse every input for its literal meaning. Comprehension does not even appear to be the simultaneous processing of literal and non-literal meanings: the literal meanings of idioms and indirect requests may not be processed at all during comprehension. Rather then supporting a notion that non-literal meanings are special and require special elaborative processing, comprehension of idioms and indirect requests can be accounted for in terms of factors that are known to be involved in language comprehension, such as phrasal meaning and frequency, or degree of conventionality"
> (Hoffman and Kemper, 1987, pp 159)

Whilst research on indirect requests and idioms does not substantiate two-process theory, this does not necessarily mean that it cannot apply to metaphors (and by extension analogy), because metaphors are often semantically anomalous with respect to their utterance context: according to Miller (1979), if the initial literal processing of an utterance yields semantically anomalous readings, this will cause the metaphor to be reconstructed as an analogy in which the non-literal comparison is explicitly stated. Again, this theory predicts longer processing time for metaphorical interpretation, since it imputes special processing which is not necessary for literal interpretation.

However, Hoffman and Kemper found in their review that experiments exploring the two-process theory of metaphor could be classified as follows:

**Fail to confirm theory:** (Harris, 1976; Harris, Lahey and Maraselek, 1980).

**Offered weak confirmation of theory:** (Petrun and Belamore, 1981; Pollio, Fabrizio, Sills and Smith, 1982). Hoffman and Kemper note that these studies fail to provide evidence that the strong predictions of two-process theory always obtain, 'and when they do obtain, the results can be explained in other ways, such as in terms of strategic biases induced by the tasks or materials' (Hoffman and Kemper, 1987, pp 167).

---

clearly any justification for these distinctions will be parasitic on some account of categorisation in much the same way as I have argued the analogy - categorisation distinction is.

**Refutation of theory:** A study by Ortony, Schallert, Reynolds and Antos (1978) provided evidence that metaphor comprehension might involve the same sort of processes as are involved in the comprehension of literal sentences (Ortony, 1980). Participants were given a target sentence (e.g. *the hens clucked noisily*), which was preceded by a context to promote either a metaphorical interpretation (about a women's debating club) or a literal interpretation (a child feeding farm animals). Ortony *et al* found that comprehension times increased (participants took longer) in relation to the amount of context available, but significantly, that these results were mirrored in the literal sentences which were acting as a control. Both with supporting context, and in its absence, literal sentences took as long to comprehend as metaphorical ones (see also Gibbs, 1984, for convergent evidence from the study of idiom processing).

However, perhaps the most interesting finding of Hoffman and Kemper's review was their discovery of the problems inherent in assessing reaction-time studies of two-process theory owing to the lack of a story other than 'categories' to justify distinguishing between what counted as a 'literal' and a 'non-literal' interpretations in all cases:

> 'There is another problem with two-process theory stemming from th[e] lack of clarity about the semantic base: the avoidance of the problem of individual differences. Basic semantic elements that are assumed to be assigned to a given word are supposed to belong to just about everyone's semantic memory, an invariant semantic hierarchy. But people have meanings, not words; differences between individuals at the level of semantic features are bound to occur.'
> (Hoffman and Kemper, 1987, pp 171)

Hoffman and Kemper's comments here may be illustrative of yet another problem inherent in relying on unanalysed intuitions in this area: the intuitions we hold about categorical behaviour may be inconsistent and contradictory. Hoffman and Kemper's remark that 'people have meanings, not words; differences between individuals at the level of semantic features are bound to occur' is revealing of a strong intuition, held both by Hoffman and Kemper and the author. It is, however, an intuition that is hard to square with definitive, essentialist accounts of categorisation: if 'knowing' a category C is simply a question of possessing a schema or essence or theory of E, then it is hard to see how individuals possessing E could differ in their understandings of C.

## 4.11 Conclusion

In this chapter I have sought to show that the evidence put forward in cognitive science research is compatible with Wittgenstein's analysis of concepts and categories; moreover, I have sought to show that the picture of categorisation that emerges from the literature cannot support the contrast with categorisation that is needed to define analogy, and also to underpin the use of 'semantics' in analogical theories.

The next chapter describe some experiments which were designed to empirically test the 'no-distinction' hypothesis.

# Chapter 5

# Probing The "Distinction" Between Cognitive Theories Of Analogy And Categorisation

## *5.1 Introduction*

The previous chapter presented the argument that the development of a psychological model of conceptual categorisation has been fatally hampered by two important, erroneous assumptions:

- Firstly, categories have often been treated as a rigid, externally imposed phenomenon;
- Secondly, researchers have concentrated upon category representation - attempting to uncover the essences of 'real' categories - rather than on the process by which categorisation judgements are made.

On the other hand, Chapter 2 showed how research in analogy has focused upon the analogical process - relating representations to one another - and the interplay between this process and the representation of analogs.

Most current research into analogy and categorisation in cognitive science has accepted a distinction between category membership and analogy based upon an essentially realist intuition. In categorical judgements, relating a new representation of an object to some kind of stored category representation, objects are felt to be similar to one another in a way in which those objects in judgements of analogical association are not. If two objects are considered to be members of a category, the classification is real; if they are considered to be analogous, it is not.

I argued above that this tacit realism has been a part of much research into analogy and metaphor. Holyoak and Thagard (1995) describe a world in which "we think we see things as they really are", and analogy is used in order to recycle our existing knowledge of the real world to formulate new bits of 'real' knowledge. Similarly, in the case of metaphor, Ortony (1979) makes a distinction between literal and non-literal similarities: 'encyclopaedias are like dictionaries' is true in a literal (real) way, whereas 'encyclopaedias are like goldmines' is only true in a metaphorical (non-real) way. Whether the notion of literal similarity might be problematic or not is barely examined, since the real problem to be addressed is metaphor. Holyoak and Thagard (1995) offer the comment "A metaphor always connects two domains in a way that goes beyond our normal category structure" (pp 217), whilst giving little indication as to what might constitute this 'normal category structure'. Analogies are defined as being distinct from categories, the nature of which are left unexamined, presumed real.

Once the difficulties of giving an account of categorisation, noted in Chapters 3 and 4, are admitted into the picture, distinctions between analogy and metaphor reliant upon a contrast with categorisation cease to distinguish at all. Analogy is consistently defined in contrast to categorisation (Clement and Gentner, 1991; Holyoak and Thagard, 1995); yet in order to make a contrast definition one needs an account of at least one of the contrasting elements. This we don't have. An analogy is defined as an associative judgement between two things that are in different categories, yet as I have sought at length to show in chapter 3 and 4, a unitary account of what constitutes an association between two things such that they are members of the same category rather than different categories is not available[1]. Moreover, the evidence reviewed in those chapters seems to indicate that according to the best accounts of categorisation, the question of whether two things are members of the same category may not be amenable to any straightforward answer. The distinction between categorisation and analogy is difficult to draw: in this chapter I shall explore empirically the hypothesis that at cognitive levels of description there may no clear distinction to be made at all.

### 5.1.1 Models Of Analogy And Categorisation

Another factor, noted earlier, which favours the abandoning of traditional distinctions between categorisation and analogy are the strong parallels which can be drawn between theories of analogy and the theoretical constraints imposed by many models of categorisation. There is widespread acceptance of the role that structure plays in

[1] Glucksberg and Keysar (1990) argue that metaphorical judgements are the *same* as categorisational judgements ("metaphors are understood as they are - as class inclusion statements", pp17). However, from the account they put forward, it is hard to see how categorisation is to illuminate metaphor, since they conclude: "The central problem is to understand categorization." ( pp 17).

category formation (Boyd, 1984; Goldstone, 1994; Keil, 1989; Medin and Ortony, 1989): analogical reasoning research directly addresses a process which reasons amongst structural networks (Falkenhainer, Forbus and Gentner, 1989; Holyoak and Thagard, 1995).

If we ignore the pre-theoretical distinction between the two processes, there are many strong similarities to note between the cognitive theories of analogy and categorisation considered so far - for instance, Gentner (1983) presents a theoretical model of analogical reasoning (reviewed in chapter 2) that shares some strikingly similar features with Medin and Ortony's (1989) account of categorisation. Indeed, it seems to capture those elements in Medin and Ortony's theory that the review in chapter 4 acknowledged were lacking in other accounts of categorisation - namely a method for capturing the important role that studies have indicated causal structure plays in conceptualisation.

### *5.1.2 Two Theories:*

Forbus, Gentner and Law (1995; pp 145-6) propose the following theoretical model of analogical reasoning:

- initial selection dependant upon surface similarity
- degree of analogical similarity is determined by deeper structures.

There is a striking parallel between this and Medin and Ortony's knowledge representations scheme for categories:

- identification procedure based upon surface features
- classification is determined by deeper structures.

In the following section, I shall examine whether these are superficial similarities, or whether they are indicative of a stronger - perhaps even an identity - relationship, between these processes in cognition.

**Story 1 - Base story**

Once there was a teacher named Mrs Jackson who wanted a salary increase. One day, the principal said that he was increasing his own salary by 20 percent. However, he said there was not enough money to give the teachers a salary increase.

When Mrs Jackson heard this she became so angry that she decided to take revenge. The next day, Mrs Jackson used gasoline to set fire to the principal's office.

*Then she went to a bar and got drunk.*

**Story 2 - Literal similarity**

Professor Rosie McGhee very much wanted a raise. One day the provost announced that he was giving himself a raise. However, he said that since money was short, no one else would get a raise this year.

After Professor McGhee heard this she became so upset that she decided to get even. One hour later, Professor McGhee blew up the administration building with dynamite.

**Story 3 - True Analogy**

McGhee was a sailor who wanted a few days of vacation on land. One day, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on the ship.

After McGhee heard this he became so upset that he decided to get revenge. Within an hour McGhee blew up the captain's cabin with dynamite.

**Figure 5.1 A**: Sample stories from Gentner, Ratterman and Forbus (1993) - the text in italic type illustrates extra structure added by Gentner at al to the base stories only.

**Story 4 - Mere appearance: (First order commonalities)**

Professor McGhee very much wanted a raise. One day she became so impatient that she used kerosene to burn down the administration building.

After the fire, the provost announced that he was giving himself a raise. However, he said that due to the fire, there was not enough money to give one to anyone else.

**Story 5 - False Analogy**

McGhee was a sailor who wanted a few days of vacation on land. One day McGhee became so impatient that he tried to blow up the captain's cabin using dynamite.

After this incident, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on board to repair the ship.

**Story 6 - Mere appearance: (Object commonalities only)**

A teacher once thought that she deserved a pay rise. She asked the principal when her rise was due. She was wearing her best suit. The principal told her that rises were decided by the governors.

**Figure 5.1 B:** Sample stories from Gentner, Ratterman and Forbus (1993)

| | BASE | LS | TA | MA | FA | MAAO |
|---|---|---|---|---|---|---|
| BASE | | Str. SM | Str. | SM | | SM |
| LS | Str. SM | | Str. | SM | | SM |
| TA | Str. | Str. | | | SM | |
| MA | SM | SM | | | Str. | SM |
| FA | | | SM | Str. | | |
| MAAO | SM | SM | | SM | | |

**Figure 5.1 C**: A schematic representation of the overlapping features in the stories from Gentner, Ratterman and Forbus (1993). SM = Surface Matches, Str. = Structure. (For details of story types, see section 5.3.2).

## 5.2 Structural Systematicity And Categorisation

Since the Gentner, Ratterman and Forbus (1993) analogy studies discussed previously did not directly address categorisation, a tacitly realist position was adopted in respect of the categories amongst which participants were to analogise. The most obvious way in which this realist assumption manifests itself is in the classification of match items (the individual stories within the "Karla the hawk" story sets (Gentner, Ratterman and Forbus 1993)). The question of the categorical status of match items is determined in advance, thus story 1 in figure 5.1 is classified as a base story, whilst story 3 is defined as its analogue. It is tacitly assumed that the two stories are members of distinct and separate categories, and that they share some kind of analogous link. Whilst the study aimed to explore a wider range of determinants of similarity, the particular correspondences determined by structural systematicity were considered to be indicative of analogous similarities (similarities *between* rather than *within* categories). These assumptions determined the predictions that Gentner *et al* made for their experiments, and the evidence they sought with which to test them.

Gentner *et al*'s study explored various criteria of similarity, and discovered that the preferred determinant of analogical similarity in participants was shared structural systematicity. As a consequence of the 'no distinction' hypothesis I predicted that if one were to use Gentner *et al*'s methods and materials to explore categorisation rather than analogy, structural systematicity might also serve as a criterion for determining category membership. Story 3 in figure 5.1 was assumed by Gentner *et al* to be an analogue of story 1. Analogues, as posited in traditional accounts of analogy, are defined in contrast to category members. If participants were to use structural systematicity as a categorisation determinant, then definitions of analogy which rely on shared structure to contrast analogy with categorisation might need some refinement. If both analogy and categorisation produce the same results, then this might imply some shared, structure based mechanism, or that one process is supervenient upon the other. Accordingly, in the following experiment, participants were presented with Gentner *et al*'s materials and asked to categorise them. Given that Gentner *et al* define the analogical mechanism in terms of structure mapping, I accordingly expected structure mapping to determine categorisation: i.e. Gentner *et al* assume that match items with only structural similarities (i.e. analogues) belong to different categories: the prediction was that they would be categorised together.

## 5.3 Experiment 1

### 5.3.1 Participants

The participants were 20 volunteers, a mixture of postgraduate and undergraduate students from the Artificial Intelligence Department at the University of Edinburgh.

### 5.3.2 Materials

The basic materials used in this study were the 20 sets of "Karla the hawk" stories (Gentner, Ratterman and Forbus, 1993).

Gentner defines the following taxonomy of similarity relationships between the stories:

• *Literal similarity* matches include both common relational structure and common object descriptions;

• *Surface matches*: based upon common object descriptions, plus some first order relations;

• *Structural similarity*, a match based upon a common system of internal relations;

• *First order matches,* where the only common feature is first order relations;

• *Object only matches,* where stories have only object matches in common.

Each set consists of a base (B), a literally similar story (LS), an analogue (TA - with only structural similarities with the base), a mere-appearance story (MA - with surface and first order commonalities with the base), a false analogy (FA - an analogue of MA), and an object only match story (OO - with only surface commonalities with the base). This allowed for a number of potential groupings according to the classification strategy adopted. Our prediction was that participants would use structural similarity as their categorical similarity determinant, putting analogues and bases into the same categories (i.e. B, LS and TA together), rather than grouping match items at the object level (i.e. grouping B, LS, MA and OO together).

The sets were modified slightly: in Gentner *et al*'s analogy research questions of the asymmetry and direction of comparisons were clearly fixed (all comparisons were in relation to the base story). Extra features (a varied mix of objects, attributes and relationships) were added to (or removed from) the base story representations (Figure 5.1, shown in italic) which did little to affect analogical similarity judgements. In categorisation judgements, aspects such as symmetry and directionality may be more fluid. As we predicted that structure would be an important determinant of categorical similarity judgements, and noting that the directionality of similarity judgements cannot be fixed in categorisation, we accordingly removed Gentner *et al*'s extra features from 65% of the story sets (G- sets: in these, for example, the base / literal similarity relationship were symmetrical), and retained the extra features (and any attendant asymmetries) in 35% of stories (G+ sets).

### 5.3.3 Procedure

Participants were randomly assigned 10 sets of 6 stories, from a total of 20 sets,[2] and asked to work through them a set at a time. Both sets and stories were presented in randomised order. For each set, they read through each story a number of times in order to familiarise themselves with its content. Participants were then asked to "Group the stories into the categories that seemed most natural and appropriate to you. These groups can range from putting every member of the story set into the same

[2]Given the sample size, the experiment concentrated on sets 1 - 10; sets 11 - 20 were used to a more limited extent to check for any marked variations in the data being produced. No such variation was noted.

group, to putting each story into a group on its own." When participants had made their categorisation decisions, they physically grouped each set of stories by pasting them onto a large sheet of paper and encircling each group in ink. Participants were then re-presented with their groupings a set at a time, asked to give each group with two or more members a simple descriptive name, and then to write a few sentences explaining what caused them to classify each named group of stories together. This task was intended to help participants focus on the naming task, and the resultant data was not analysed.

| Classification Criterion | | | | % of Total |
|---|---|---|---|---|
| *Systematic network of relations in common* - Type 1 | | | | 79.5 % |
| 1 B LS TA | 2 FA MA | 3 OO | | |
| *Systematic network of relations in common* - Type 2 *(Base classified separately)* | | | | 8 % |
| 1 LS TA | 2 FA MA | 3 B | 4 OO | |
| *First order relations in common* - Type 3 | | | | 4 % |
| 1 B LS TA FA MA 2 OO | | | | |
| *Only object similarities in common* - Types 6 & 7 | | | | 5 % |
| 1 MA LS B OO | 2 FA TA | | | |
| 1 B OO | 2 LS MA 3 TA FA | | | |
| *No classification possible* - Types 4, 5, 8 , 9 10 | | | | 3.5 % |
| 1 B LS MA | 2 FA TA 3. OO | | | |
| 1 B TA | 2 FA MA | 3 LS | 4 OO | |
| 1 B MA FA | 2 LS TA | 3 OO | | |
| 1 B LS TA OO | 2 FA MA | | | |
| 1 FA B TA MA | 2 LS OO | | | |

**Figure 5.2:** Output patterns from the categorisation task, showing the groups formed and criteria established. The stories are labelled according to how they fit Gentner's taxonomy of similarity (defined above): B = Base; LS = Literal Similarity; TA = True Analogy; FA = False Analogy; MA = Mere Appearance; OO = Object Only match.

### 5.3.4 Results

Each story set assigned to a participant was analysed to determine the groups formed. The pattern of groupings which emerged fell broadly into 5 classes (figure 5.2). Gentner *et al*'s taxonomy of similarities could account for 96.5% of within group similarities observed. Of these, in 5% of cases the stories were grouped according to types 6 and 7. The only similarities across groupings in these types are that the stories in the individual groups had only objects in common. In 4% the stories were classified according to type 3, where the across grouping similarity was shared first order relations. In 79.5% of cases participants grouped using type 1. Here the only similarity across groupings was a network of systematic causal relations. The full output and incidence of the types is given in Table 5.1.

8% of groupings were according to type 2, where the base was put into a category on its own, with the only similarity across other groupings being shared structure. This type was only found once amongst those sets from which Gentner *et al*'s extra features had been removed (0.5% of G- sets; figure 5.3). The G+ sets, those with added features in the base, were sets 5; 7; 10; 12; 15; 17; and 20. Of these: in set 5 and set 20 the extra features involved higher order relations; in sets 7; 10; and 15 they involved first order relations; and in sets 12 and 17 the extra features were objects. 20% of these sets were classified as type 2, with the bulk of these classifications being in the sets with extra higher order relations (figure 5.4).

| | **Story Set Type** | | |
|---|---|---|---|
| **Grouping Type** | *G+* | *G-* | *% of total* |
| *Type 1* | 68 % | 86.5% | 79.5% |
| *Type 2* | 20 % | 0.5% | 8% |
| *Other* | 12% | 13% | 12.5% |

**Figure 5.3:** Classification strategies according to set type.

| | *Higher-order relations* Sets 5 | 20 | *Objects only* Sets 12 | 17 | *1st order relations* Sets 7 | 10 | 15 |
|---|---|---|---|---|---|---|---|
| Type 1 | 7 | 2 | 7 | 7 | 11 | 9 | 5 |
| Type 2 | 9 | 2 | 1 | | 2 | | |
| Type 3 | 1 | 1 | | | 1 | 2 | |
| Types 6 & 7 | 2 | | | | | | |
| Types 4, 5, 8, 9 & 10 | | | | | 1 | | |
| Totals | 19 | 5 | 8 | 7 | 14 | 11 | 6 |

**Figure 5.4:** Classification data for the G+ sets.

**Story Sets**

| Participants | 1 | 2 | 3 | 4 | 5+ | 6 | 7+ | 8 | 9 | 10+ | 11 | 12+ | 13 | 14 | 15+ | 16 | 17+ | 18 | 19 | 20+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A** | | | | | **1** | **1** | | **1** | **1** | **1** | | | **1** | **1** | **1** | | **1** | **1** | | |
| **B** | **3** | **3** | **1** | **1** | **2** | **1** | **1** | **1** | **1** | **1** | | | | | | | | | | |
| **C** | **1** | **1** | **1** | **1** | **2** | **1** | **1** | **1** | **1** | **1** | | | | | | | | | | |
| **D** | **1** | **1** | **1** | **1** | **2** | | **1** | | | | **1** | **2** | | | | | | | **1** | **2** |
| **E** | **1** | **1** | **1** | **1** | **2** | **1** | **2** | **1** | **1** | **3** | | | | | | | | | | |
| **F** | **4** | **1** | **1** | **1** | **2** | **1** | **1** | | | | **1** | **1** | | | | | | **5** | | |
| **G** | | **8** | | | **6** | **7** | | **1** | **1** | | | | **1** | **6** | **1** | | **1** | **1** | | |
| **H** | **5** | **1** | **1** | **1** | **2** | | **1** | | | | **1** | **1** | | | | | | | **1** | **1** |
| **I** | **1** | **1** | **1** | **1** | **1** | | **1** | | | | **1** | **1** | | | | | | | **1** | **2** |
| **J** | **1** | | **1** | | **8** | **1** | | **1** | **1** | | **1** | | | | **1** | | **1** | **1** | | |
| **K** | | | | | | | **1** | **1** | **1** | **1** | | **1** | **1** | **1** | | **1** | **1** | | | **1** |
| **L** | **7** | **1** | **1** | **1** | **2** | **1** | **1** | **1** | **1** | **1** | | | | | | | | | | |
| **M** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | | | | **1** | **1** | | | | | | **1** | | |
| **N** | **1** | **1** | **1** | **1** | **5** | **2** | **2** | | | | **1** | **1** | | | | | | **1** | | |
| **O** | | | | | **1** | **1** | | **1** | **1** | **1** | | | **1** | **1** | **1** | | **1** | | **1** | |
| **P** | **6** | **3** | **1** | **1** | **1** | | **3** | | | | **7** | **1** | | | | | | | **1** | **3** |
| **Q** | | | | | **2** | **1** | | **1** | **1** | **1** | | | **2** | **1** | **1** | | **1** | **1** | | |
| **R** | **1** | **10** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | | | | | | | | | | |
| **S** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | | | | | | | | | | |
| **T** | | | | | **3** | **7** | | **9** | **1** | **3** | | | **7** | | **7** | **7** | **1** | **1** | | |

**Table 5.1:** Output incidence of participant groupings. Each participant was given 10 story sets (each row represents one participant): the type of grouping is indicated by the type number in the story set column (see also figure 5.2). (Participant T produced some rather strange results: this was explained by examining the reasons T gave for her groupings, in which she explained that she was exploring a different heuristic for each story set.)

## 5.4 Discussion

Experiment 1 examined the hypothesis that mechanisms normally considered to be analogical could in fact support categorisation tasks. The most important finding here is the role that shared structure plays in classification judgements in this task. 79.5% of the groupings formed by our participants had only shared systematic structure (traditionally defined as analogy) as a common feature amongst members of the categories formed. In contrast, only 5% of groupings produced had common object descriptions as the common similarity within categories.

The argument put forward above, that abandoning the current de facto distinction between categorical and analogical associations of objects might benefit the understanding of both, has two main parts: first, that the standard distinctions (Clement and Gentner, 1991; Holyoak and Thagard, 1995) between analogy and categorisation fail to actually distinguish between them, leading to problems in saying quite what analogy is supposed to comprise; and secondly, that by removing the distinction, understanding of the factors which govern mappings between representations that have been gleaned from analogy research might help illuminate categorisation questions. These results provide some evidence in support of the latter idea that structures, and more pertinently Gentner's structural systematicity, rather than simple features, can act as the key to categorical similarity.

This argument is supported not only by the proportion of categorisations that were determined by commonalities between internal structures in the stories, but also by the effects of added structure in the G+ sets where the added structure was a higher-order structure. These might at first appear to present a problem for an attempt to use a structure mapping analysis to model these categorisation judgements. In these cases, Gentner's base stories were put into separate categories from stories to which they were supposed to be literally similar, which were in turn categorised alongside their supposed analogues (both of which were supposed to share structures with the base).

These results can be attributed to the effects of directionality and symmetry upon similarity judgements. Whilst Gentner, Ratterman and Forbus (1993) found that participants judged literally similar (LS) stories to be very similar to bases, and analogues less so, they did not consider the effect of reversing the directionality and symmetry of the comparisons, for example comparing the base and analogue stories' similarity to the LS. Neither did they consider the judging of cumulative similarity, where dissimilarities are also taken into account. During this process, the structural dissimilarities of the base versus the LS and analogue appear from our results to be clearly relevant, whereas the object differences of the analogue versus the LS and base do not. This maximisation of important similarities (i.e. structure matches) relative to lesser dissimilarities (i.e. object matches) amongst groupings appears to play a crucial role in categorisation in this study.

Whilst it might be argued that all I have shown here is that participants will form categories of analogies, such an interpretation (in so far as one can make sense of it) does not affect the argument that it is common systematic structure that determines the content of these categories.

## *5.5 Using Surface Versus Structural Recall Biases To Probe Conceptual Storage*

A major problem in much research into categorisation, highlighted in the review in chapter 4 above, is that experimental results have rarely - if ever - directly indicated anything about conceptual representation. Often, if not always, it is difficult to determine whether particular results stem from stored information regarding concepts (e.g. propositional or imaginal information) or from the processes that operate in invoking a particular concept (c.f. Smith and Medin, 1981). The vast majority of theories reviewed in chapter 4, assume a 'straightforward' unitary representation, the exception being granular instance-based approaches to categorisation, and perhaps explanation-based approaches (although the lack of any specific formulation of an explanation (or theory) based model of categorisation makes it impossible to deduce the kind of stored representations such a theory would entail).

In the light of this, and the theoretical analysis presented in chapter 4, direct evidence regarding the nature of the storage of a concept, or concepts, may have important implications for the way categories are viewed: especially if that evidence fails to support a unitary-representation account.

## *5.6 A Process Model*

In experiment 1 I have examined an intuitive distinction - held by most psychological researchers - between analogy and metaphor on one hand, and categorisation on the other. I have argued that although one might ordinarily distinguish between category membership and analogy according to realist terms, there are good reasons for abandoning this distinction at a theoretical level when we focus upon cognitive processes.

### *5.6.1 Representation And Process*

Studies by Gentner, Ratterman and Forbus (1993) (discussed in chapter 2 above) explored the factors determining how items are accessed, i.e. how representations are selected in order to allow similarity mapping to take place. These have shown that access relies primarily upon surface attribute (or object attribute) matches, and they propose that the process underlying judgements of similarity can be decomposed into two sub-processes:

- *Accessing* a similar (*base*) situation from memory, based primarily on surface similarity
- Creating a *mapping* from base to target using structural commonalities (SMT).

I shall explore whether this process model can offer a solution to the difficulties, mentioned above, of determining whether particular effects result from stored representations or from the processes that operate in invoking representations. The following experiment was designed to use this detailed model of the processes that determine the retrieval and mapping (and classifying) of representations to empirically probe the nature of conceptual storage.

**Story 1 - Base story**

Joseph was a millionaire who hired a chauffeur to drive his Rolls-Royce. He used to brag to his wife that he would never be late for his conferences since he had hired a chauffeur.
One morning when he was in a great hurry, he went to find the chauffeur. But he was asleep. He thought his services would not be needed that day.
Thus Joseph was very late for his conference after all. To make sure this would not happen again, Joseph hired a second chauffeur.

**Story 2 - Literal similarity**

Alexander was a wealthy man who employed a driver for his limousine. He liked to boast to his spouse that with his driver he would always be on time for his meetings.
One day when he was in a rush, he went to find the driver. But he was taking a nap. The driver thought it was his day off. Thus Alexander ended up missing his meeting. But to make sure he would not be late again, Alexander hired a second driver.

**Story 3 - True Analogy**

Alexander was a man who lived with his wife in his house a long time ago. He liked to boast to a friend that with his wife at home we would always eat well. One day, when he was very hungry, Alexander went home to his wife. But she thought Alexander would be eating someplace else so she had only prepared enough for herself and their baby. Thus Alexander went without dinner. But to make sure he would not go hungry again, Alexander married a second wife.

**Figure 5.5**: A classified (B,LS,TA) story set (Gentner, *et al*, 1993) from experiment 1, as used in experiment 2

## 5.7 Experiment 2

In the course of experiment 1, participants were asked to give each of their classes a name that was meaningful to them. Because of the particular nature of that task, this involved participants developing (and learning) 'categories' that contained some items that had only structural relations in common. By examining the attributes they could recall that were associated with that name, I aimed to use Gentner *et al*'s findings about systematic structures vs. attributes to determine the representation associated with the name. If participants stored some kind of abstracted prototype - i.e. a unitary representation of their category - we would expect that the attributes associated with the most prototypical stories would be most readily retrieved from memory, with attributes recalled insofar as they were relevant to the prototype (perhaps along the frequency lines one might expect from the analysis in table 5.2).

This was not expected to happen. There is too much empirical and theoretical work that cannot be accommodated by a unitary representation account (see chapter 4). The hypothesis here was that participants would not have abstracted a unitary representation from the stimuli that they had classed together, but would instead store a number of representations that they associated with it, polarising between those class members that shared no surface attributes. I predicted that introducing a class name out of context would make it equally likely that any stored representation associated with that name would be recalled, and that the initially recalled representation would drive further recall. Since the model of recall employed here is feature driven, I expected results to polarise, with those representations with no surface attributes in common leading to minimal recall of one another.

### 5.7.1 Participants

The participants were 20 volunteers, a mixture of postgraduate and undergraduate students from the Artificial Intelligence Department at the University of Edinburgh.

| Set | B-LS | LS-SS | SS-B |
|---|---|---|---|
| 1 | 100 | 90 | 85 |
| 2 | 100 | 100 | 85 |
| 3 | 100 | 65 | 45 |
| 4 | 100 | 55 | 45 |
| 5 | 100 | 85 | 55 |
| 6 | 100 | 85 | 75 |
| 7 | 100 | 75 | 60 |
| 8 | 100 | 70 | 45 |
| 9 | 100 | 85 | 65 |
| 10 | 100 | 75 | 60 |
| 11 | 100 | 95 | 75 |
| 12 | 100 | 75 | 60 |
| 13 | 100 | 75 | 50 |
| 14 | 100 | 95 | 70 |
| 15 | 100 | 50 | 35 |
| 16 | 100 | 95 | 75 |
| 17 | 95 | 100 | 70 |
| 18 | 100 | 80 | 55 |
| 19 | 100 | 80 | 60 |
| 20 | 100 | 100 | 60 |

**Table 5.2:** Object attribute (surface feature) similarity ratios between stories by set.

### *5.7.2 Materials*

Materials were the classified sets of "Karla the hawk" stories (Gentner, Ratterman & Forbus, 1993; see figure 5.5 for an example) produced by the participants in experiment 1. The following taxonomy of similarity relationships between the stories can be defined:

- *Literal similarity* matches include both common relational structure and common object descriptions;
- *Surface matches*: based upon common object descriptions, plus some first order relations;
- *Structural similarity*, a match based upon a common system of internal relations;

The original sets consist of a base (B), a story literally similar to it (LS), one structurally similar to it (SS - i.e. with no object attributes in common with the base, although SS stories did share some object attributes with LS). In order to determine the relative effects of object versus structural matches in this experiment, the "Karla the Hawk" stories were analysed and rated to determine the level of attribute commonalities between the individual stories in each set.

| Correspondence | Example | Value |
|---|---|---|
| 1 -1 map | man, man | 5 pts |
| strong map | street, road | 4 pts |
| 'analogical' association | conned, robbed<br>fireman, paramedic | 3 pts |
| weak 'analogy' | enlisted, begged<br>fireman, nurse<br>disappeared, shattered | 2 pts |
| weak association | Fred, Mary (both names)<br>fireman, accountant (both jobs) | 1 pt |

**Figure 5.6:** Classification criteria for determining surface similarity.

Two raters gave a numeric value to each of a range of possible surface attribute correspondences between stories (see figure 5.6 for details), and then individual attribute correspondences were totalled and averaged between the two raters in order to determine the overall correspondences between stories (table 5.2). Consistency between raters was 82.5%. Differences between raters were resolved by discussion.

### 5.7.3 Procedure

The classes produced by participants in experiment 1 were returned to them. During the classification task, participants were asked to give each of the classes they produced a name that 'would be meaningful to them later'. After finishing the classification task, participants were given a 5 minute break, and then undertook a 20 minute diversionary task (searching for post-codes from a directory) before being given another 5 minute break. Participants were then presented with a sub-set (usually 4) of the names they had assigned to classes during the classification task, and asked to 'write down what you can remember about the various features (or you may like to see them as attributes) of each of the scenarios associated with each name. E.g. you may have had a scenario about a door that needed varnishing. Features, or attributes, associated with such a scenario would be "door" and "varnishing"'. Participants were given 10 minutes to complete the task.

### 5.7.4 Results

The 20 participants yielded a total of 70 recall episodes.

#### 5.7.4.1 Scoring

The recalled features were evaluated by two judges using the same scale that was used to evaluate feature correspondences between stories (see figure 5.7). The total attribute recall for each set was calculated and averaged between raters. As in the rating of story commonalities, differences between rater were resolved by discussion.

#### 5.7.4.2 Individual Story Recall

B was best recalled for 37% of all sets, SS in 30% of cases and LS in 33%; as predicted, there was no significant bias towards recalling any particular type of story. However, when we looked at the pattern of recollection, irrespective of the particular stories each participant had recalled, there were significances in the quality of recall between the best recalled, and the next best, and the next-best and the worst recalled stories. Participants tended to clearly recall one story better (*70 cases*, $M$ = 19.75) than the next ($M$ = 14.74), (within groups $t(69) = 8.846$, $p < 0.0001$), and then these next best recalled stories better than the worst ($M$ = 11.1), $t(69) = 11.802$ $p < 0.0001$

The individual recall orderings by story type are given below:

#### 5.7.4.3 Base (B)

In cases where B was the most recalled story in terms of features (26 cases, $M$ = 18.44), there was a significant difference in the quality of recall over the next best recalled story set LS (M = 14.83), $t(25) = 4.434$ $p < 0.0001$, which in turn was recalled significantly more than SS (M = 10.15), $t(25) = 3.77$ $p < 0.0001$.

#### 5.7.4.4 Literal Similarity (LS)

When LS was the best recalled story (M = 20.78), next best was B (23 cases, $M$ = 15.85), $t(22) = 5.288$ $p < 0.0001$, with SS recalled least well ($M$ = 11.96), $t(22)$ = 2.095 $p < 0.05$.

#### *5.7.4.5 Structural Similarity (SS)*

In cases where SS stories were qualitatively best recalled (21 cases, $M = 20.12$), the next best recalled group was LS, ($M = 13.43$), $t(20) = 5.886$ $p = < 0.0001$, and the least recalled group was B ($M = 11.36$), although the between group difference between LS and B was not significant ($t(20) = 1.901$ $p < 0.072$); the difference in recall quality between SS and B was still significant ($t(20) = 6.609$ $p < 0.0001$).

| **B Best Recalled** | Next Best - LS | Next Best - SS |
|---|---|---|
| 26 cases;<br>Mean = 18.44 | Mean = 14.83 | Mean = 10.15 |

| **LS Best Recalled** | Next Best - B | Next Best - SS |
|---|---|---|
| 23 cases;<br>Mean = 20.78 | Mean = 15.85 | Mean = 11.96 |

| **SS Best Recalled** | Next Best - LS | Next Best - B |
|---|---|---|
| 21 cases;<br>Mean = 20.12 | Mean = 13.43 | Mean = 11.36 |

**Table 5.7:** Mean recall orderings by story type

## *5.8 Discussion*

The experiment produced little evidence to support the hypothesis that our participants had abstracted and stored schemas from the groups they had classified, despite the fact that a shared structural schema was the basis of participants' original classification decisions. If some version of a stored prototype theory were true, we would have expected a majority of LS features to be recalled in most instances. In fact, B features were most often recalled, though not significantly: the trend favoured a random distribution. Another result that might also favour prototype theory would have been a situation where all the stories were recalled with much the same frequency, i.e. LS=B=SS, since such a result could be a product of the strong feature commonalities between the LS stories and members of both of the other story types. However, there was a significant trend for participants to recall one story more than another, and the next best story more than the least recalled story.

If participants randomly recalled an individual instance of a class then one would expect from Gentner *et al*'s similarity recall findings that the attributes of this member should influence which other story they might recall from the class: if B is recalled, recall of a B story should prompt recall of an LS story rather than a a SS story, as B shares more surface attributes with LS than SS (B and LS share 10 surface attributes to every 6 shared by B and SS), and Gentner *et al*'s findings were that surface attributes rather than shared structure promote recall: where B features were best recalled this pattern emerged throughout our study. As predicted, the result of LS stories sharing a higher percentage of surface attributes with B than SS shared with B, results in a situation where recall of B led to a significantly higher quality of recall for LS recall than SS. The results where LS stories were most strongly recalled also supported this analysis, with SS recall prompting significantly better recall of B attributes than SS attributes (even though LS shared more surface features with SS than SS did with B).

Indicative of the fact that SS shared fewer surface commonalities with the other story types, SS-LS attribute commonalities were much weaker than B-LS (see table 5.2), and in specific sets little greater than B-SS. Good SS recall did not produce a bias towards LS or B as the next-best recalled story type; although our results showed some tendency towards SS prompting LS over B, it was not significant (see also table 5.3).

The hypothesis that participants would recall stories individually from memory is further supported by the nature of participant's recollection. Irrespective of the particular stories each participant had recalled, there were significances in the quality of recall between the best recalled, and the next best, and the next-best and the worst recalled stories.

If one applies Gentner, *et al*'s (1993) persuasive analysis of the influence of surface vs. deep structure on recall to these results, it would appear that being presented with a class name in no particular context caused the participants to randomly recall one of the examples associated with that name, and then use that example as the stimulus for recalling other class members. On this evidence, it would appear that participants had stored class examples along with a cue - the class name - rather than any generalisation of the class itself.

These findings suggest that at least some concepts can be stored as multiple-representations, as opposed to stemming directly from the particulars of unitary conceptual schemas, as has often been assumed.

## 5.9 Can One Distinguish Between Cognitive Theories Of Analogy And Categorisation?

So far I have reviewed and presented much evidence, both theoretical and empirical, which casts doubt on a two-process account of literal (categorical) versus non-literal (analogical or metaphorical) reasoning (two-process in the sense that the computation of non-literal meaning is supposed to be a separate process that is - in some way - parasitic on literal processing).

Despite this weight of evidence contra the two-process account - and the concomitant lack of evidence for it - suggestions that the two processes-account be rejected are still met with some incredulity. French (1995) describes the experience of suggesting to an academic audience that an upturned orange-crate, when covered with a cloth and laid out with a picnic, might *really* be described as a table. This met with the swift response, “An orange crate is an orange crate is an orange crate.” The attachment to pre-theoretical intuitions is a strong one, even amongst those who seek to explore them.

The belief that an orange crate is an orange crate is an orange crate holds great sway. Indeed, such is the two-process account’s entrenchment in ordinary, pre-theoretical understandings of the world that, quite understandably, a more than usual quantity of counter-evidence may be required if it is to be abandoned at a theoretical level. And, whilst the study of experiments 1 and 2 might apply to classifying *stories*, it provides little evidence that this classification of stories can be generalised to other categorisation behaviour.

**Base**

A Karla is a novel type of cooking pot, used by the Timuni in Alnata.
The structure of the Karla is designed in order to reduce the heat inside, and therefore prevents the food getting burned in the scorching cooking fires.
Water is poured into a layer of the Karla during cooking, which cools the food.

**Literal Similarity**

The Valkri is a special kind of frying pan, used by the Jalpeni in Frodon.
The Valkri is created in such a way as to be able to reduce heat, thereby preventing meat being getting burned when using the extreme temperatures of the cooking fires.
A liquid is poured into the layers of the frying pan when cooking, which cools the temperature of the meat.

**Structural Similarity Only**

The Vubu is a special wall built by the Jakar tribesmen in Frodon.
The Vubu is built in such a way as to be able to reduce the heat within it, thereby preventing the Jakar from sweating too much in the extreme temperatures of the midday sun.
A liquid is pumped through the Vubu, which cools the stone and therefore prevents the Jakar within the walls from getting too hot.

**Figure 5.8 A:** Example of an object description set used in the study.

ppearance

ople of Frodon use a special type of frying pan, known as the

lkri is designed in order to allow it to be handled by children, as
be difficult.
dle is designed with a special U-shape, which enables it to be held
ple with small hands.

rally Similar to MA

kar tribesmen of Frodon have built a special wall known as the

ibu's stone gates can be opened by elderly people, despite their
weight.
s set in the wall incorporate springs, which allow weaker people to
e gates.

Only

type of cooking pot, called the Karla, is used by the people of
Karlas can be purchased in a range of colours. Food cooked in a
stes great.

5.8 B: Example of an object description set used in the study. Experiment 1
ed that shared structural systematicity (Gentner, 1983) - typically defined as
y - was the key determinant in participants' categorising in their study. In the
f this finding, The hypothesis was that because structural commonalities in the
escription sets ran across the 'normal' categories embodied in the set's object
ptions, these 'normal' category boundaries would be ignored as participants
categorised objects according to their shared structural systematicity.

| LS | SSO |
|---|---|
| Str. SM | Str. |
| | Str. |
| Str. | |
| SM | |
| | SM |
| SM | |

natic representation of the
ns. SM = Surface Matches.
for Object Description t

## l Systematicity

ment[3] was designed to
ent 1 and 2: that a single-
hould be taken of the
iment 1 utilised the 'Karl
designed as materials for
typically episodic stories,
features of each story set
se findings, new sets of
s, but which rather than b
in, with a controlled varia
member; see Figure 5.8).

---

arrington and Toby Lingstadt.
who actually ran these experim

| Set Member Description | | | | | | |
|---|---|---|---|---|---|---|
| **Object Sets** | **B** | **LS** | **SSO** | **MA** | **SMAO** | **OO** |
| **SET 1** | plant | plant | tribe | plant | tribe | plant |
| **SET 2** | country ruler | country ruler | leading animal | country ruler | leading animal | country ruler |
| **SET 3** | board game | board game | field game | board game | field game | board game |
| **SET 4** | animal | animal | tribe | animal | tribe | animal |
| **SET 5** | cooking utensil | cooking utensil | wall | cooking utensil | wall | cooking utensil |
| **SET 6** | animal | animal | priest | animal | priest | animal |
| **SET 7** | chant | chant | game | chant | game | chant |
| **SET 8** | food | food | drink | food | drink | food |

**Figure 5.9:** 'Normal' categories embodied in each object description set.

The principal reasoning behind this is straightforward: typical 'natural' categories - the kind of categories found and used in human societies - tend to concern objects, and other, more regular 'things in the world', rather than stories (Rosch, 1978). Classifying objects is more akin to ordinary categorisation than classifying stories (though a set of descriptions of 'rituals' was also included to capture the fact ordinary categories reflect a good deal more than just physical regularities). Thus, the resulting object descriptions fell neatly and clearly into 'normal' categories (see Figure 5.9).

## 5.11 Experiment 3

### 5.11.1 Participants

20 volunteer participants participated in this experiment. The participants were a mixture of Artificial Intelligence and Psychology students from the University of Edinburgh.

### 5.11.2 Materials

The basic materials used for this study were 8 sets of 'Karla the Pot' novel object descriptions (see Figure 5.8 for examples). These were descriptions of objects created to replicate the framework used by Gentner, Ratterman and Forbus (1993) in the creation of the "Karla the Hawk" stories.

By analogy with Gentner *et al*'s taxonomy of similarity, used in experiments 1 and 2, the following taxonomy of similarity relationships was defined between the novel "object descriptions":

- *"Literal similarity"* matches include both common relational structure and common object attributes;

- *"Surface matches"* are based upon common object attributes, plus some first order relations;

- *"Structural similarity"* matches are based upon a common system of internal relations;

- *"First order"* matches only have first order relations as a common feature;

- *"Object only"* matches only have object matches in common between the object descriptions.

Each of the 'Karla the Pot' sets consists of a base (B), a literally similar object description (LS), an object description that shared the same structure as the base, but no object attributes (SSO)[4], a mere-appearance object description, with surface and first order commonalities with the base (MA), an object description which shared structure with the MA, and object attributes with the SSO (SMAO),[5] and an object only match object description, with only surface attribute commonalities with the base

[4] The same as TA in experiment 1.

[5] The same as FA in experiment 1.

(OO). This allowed for a number of potential groupings to be formed, according to the classification strategy participants adopted.

The prediction examined was whether participants would again use structural similarity as their categorical similarity determinant, putting analogues and bases into the same categories (i.e. B, LS and SSO together), rather than grouping match items at the object level (i.e. grouping B, LS, MA and OO together; which also equated to existing category membership), despite the fact that we were using novel object descriptions which embodied existing categories rather than Gentner's (1993) relatively 'category-neutral' stories.

### 5.11.3 Procedure

Each participant was presented with eight envelopes, each containing a different set of six novel object descriptions, and was asked to work through them one set at a time. Sets were presented in random order, as were the object descriptions within them.

Participants were instructed to read through the object descriptions within a set several times, until they felt familiar with their contents. They were then asked to put the objects together into groups, grouping the things that fitted most naturally together in their judgement. Groupings could range from putting all descriptions into the same group to having them all in separate groups as well as all variations in between.

When the categorisation groupings had been decided on, the participant pasted them onto a large sheet of blank paper and then circled each grouping using a marker pen.

Once all eight sets had been divided into groups using this procedure, participants were re-presented with their groupings a set at a time, and were asked to give any group containing two or more members a simple descriptive name.

The experiment took around an hour to complete.

## 5.12 Results

For every object description set, the groups formed by each participant's classifications were analysed (with the results displayed in Table 5.3). Groupings which emerged fell into a number of classes, a taxonomy of which is presented in figure 5.4. Similarities across groupings (i.e. within groups similarities) which could be determined according to Gentner *et al*'s taxonomy were identified in 80% of groupings (in Types 1, 3, 4, 5 and 6).

| Subjects | Object Description Sets 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total No. of Type 1 Responses |
|---|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | 1 | 5 | 1 | 7 | 1 | 7 | 5 |
| B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| C | 1 | - | 1 | 1 | 1 | 3 | 1 | 1 | 6 |
| D | - | 1 | 1 | 1 | 1 | 8 | - | 1 | 5 |
| E | 1 | 1 | 1 | 1 | 4 | - | 1 | 1 | 6 |
| F | 1 | 1 | 1 | 1 | 1 | 1 | - | 7 | 6 |
| G | 1 | - | 1 | 1 | 1 | - | 1 | - | 5 |
| H | - | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| I | 1 | 1 | 1 | 1 | 6 | - | 1 | 1 | 6 |
| J | 1 | 1 | 2 | 1 | 1 | - | - | 1 | 5 |
| K | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| L | 1 | 1 | 1 | 1 | 1 | - | 1 | 1 | 7 |
| M | 1 | 6 | 1 | 6 | 1 | 1 | 1 | 8 | 5 |
| N | 8 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 6 |
| O | 1 | 1 | 1 | 1 | 1 | 1 | 6 | 6 | 6 |
| P | 1 | 6 | 1 | 1 | 1 | 7 | 1 | 7 | 5 |
| Q | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| R | 1 | 7 | 1 | - | 1 | - | 6 | 7 | 3 |
| S | 9 | 1 | 1 | 6 | 1 | - | 9 | - | 3 |
| T | 1 | 4 | 3 | 1 | 3 | - | 4 | 9 | 2 |

**Table 5.3:** Results for grouping patterns. Each participant was given 8 sets of object descriptions (each row represents one participant; each column an object description set): the type of grouping is indicated by the type number in the object description set column (see also Table 5.4).

The most common grouping pattern used was of Type 1 (groups divided into: 1. B-LS-SSO: 2. SMAO-MA: 3. OO, using a network of systematic causal relations), which was used for 70% of all object description sets.

Object description sets were grouped according to Types 4 and 5 in 3.1% of cases. The only similarity across groupings of these types is that the object descriptions in each group had only objects in common.

Groupings which occurred due to participants using common first order relations (those of Type 3) occurred in 1.9% of cases.

Other groupings worth mentioning were Types 7 and 8, in which the structured object descriptions were grouped according to a determinable pattern, (structure for Type 7, 4.4%) and object attributes (Type 8, 1.9%), but the OO descriptions were assigned according to features in Type 7 (where one would expect a separate grouping), and grouped separately in Type 8 (grouped with descriptions containing similar object attributes expected).

Only 0.6% (one occurrence) of groupings were of Type 2, where the base was put into a category of its own, with shared structure being the only similarity across groupings.

In 11.2% of groupings it was impossible to determine an overall criterion for determining the pattern produced; each of these groupings had only a single occurrence.

## 5.13 Discussion

This study further examined the 'no distinction' hypothesis: that categorisation judgements in humans can be determined more by shared structural systematicity than by shared object attributes (surface features), between the objects/ things/ rituals to be classified. The results show considerable evidence to support this hypothesis: 70% of the groupings were made in this way (had participants grouped randomly, mathematical combinatorics yield 213 possible groupings of the materials). In a further 10% of groupings (Types 2, 6 and 7), shared structure was clearly the criterion determining the participants' overall groupings, although a single object description was classified unaccountably.

| **Classification Criterion** | **Number** | **% of Total** |
|---|---|---|
| Systematic network of relations in common - Type 1 | | |
| 1 B LS SSO  2 SMAO MA  3 OO | 112 | 70% |
| Systematic network of relations in common - Type 2 (see experiment 1) (Base classified separately) | | |
| 1 LS SSO  2 SMAO MA  3 B  4 OO | 1 | 0.6% |
| First order relations in common - Type 3 | | |
| 1 B LS SSO SMAO MA  2 OO | 3 | 1.9% |
| Only object similarities in common Types 4 & 5 | | |
| 1 MA LS B OO  2 SSO SMAO | 4 | 2.5% |
| 1 B OO  2 LS MA  3 SSO SMAO | 1 | 0.6% |
| Largely systematic network of relations in common  Type 6 | | |
| 1 B LS  2 SSO  3 MA SMAO 4 00 | 8 | 5% |
| OO 'Problems' - Structure based - Type 7 | | |
| Object attribute based - Type 8 | | |
| 1 B LS SSO OO  2 MA SMAO | 7 | 4.4% |
| 1 B LS MA  2 SSO SMAO  3 OO | 3 | 1.9% |
| Type 9 and others - No clear pattern | | |
| 1 B LS  2 SSO  3 MA  4 SMAO  5 OO | 3 | 1.9% |
| Others | 18 | 11.2% |

**Table 5.4:** Output patterns from the categorisation task, showing the groups formed and criteria established. The object descriptions are labelled according to Gentner's taxonomy of similarity (defined above): B = Base; LS = Literal Similarity; SSO = Structural Similarity Only; SMAO = Structural Similarity with MA and Object Similarity with SSO; MA = Mere Appearance; OO = Object Only match.

One interesting effect from the experiment 1 that - intentionally - was not replicated in this experiment, was the production of a large number of Type 2 groupings. In experiment 1 an extra structure (inserted by Gentner *et al* as part of their analogy study) was left in a subset of the base stories presented to participants. These base stories with extra structure then tended to be grouped singularly (see Type 2 in Table 5.4, below). Since the 'Karla the Pot' materials did not contain any extra structures in the Base, I did not expect significant numbers of Type 2 stories to be produced, and in the event, only 0.6% of groupings (1 out of 180) resulted in a Type 2 pattern, where the base was classified singularly in an otherwise structurally determined grouping pattern.

Groupings that appeared to be formed on the basis of shared surface attributes only amounted to 3.1% of the total (Types 4 & 5). To these could be added another 1.9% of groupings (Type 8) in which shared features determined the overall groupings, although the OO object description - distinctive due to its complete lack of any systematic structure - was classified separately.

Of those object descriptions classified according to shared object attributes, only 2.5% (Type 4 groupings) reflect the 'normal' categories shown in figure 5.9.

Clearly, once again structure appears to be the key determinant of participants' classifications in this study. As Murphy and Medin (1985) note, categorisation models have tended to concentrate on object descriptions, making use of very representationally-simple attribute-value lists. In contrast, analogy research has examined relationships between highly structured representations (considering the influence of attributes, relations and higher-order relations in judgements of similarity). The evidence of this study would appear to support the idea put forward in chapter 4, that more notice needs to be taken of the kinds of representations used, and the effects these produce, in categorisation studies (see also Medin, Goldstone, and Gentner, 1993).

## 5.14 Summary

Experiment 1 explored the hypothesis that mechanisms normally considered to be analogical could in fact support categorisation tasks. Participants were given Gentner *et al*'s analogy materials and instead of being given analogy tasks, they were asked to categorise these materials. Participants' judgements of categorical similarity between the stimuli were indistinguishable from the judgements of analogical similarity reported by Gentner, Ratterman and Forbus (1993).

Experiment 2 provides evidence in support of the hypothesis that it is possible for participants to name a class whose membership is determined entirely by shared structure, and then can retrieve information regarding its members without appearing to abstract a common schema definitive of that class. Such evidence for a non-unitary account of category representation should not come as a surprise: the literature is filled with material that casts doubt on the plausibility of unitary accounts of concept representation.

Finally, the results of experiment 3 add further support the broader finding of experiment 1, which indicated that the processes underlying analogy and categorisation are not as distinct as is usually assumed. The results of both experiments show shared structural systematicity (Gentner, 1983) as the main process underlying categorisation judgements in the particular experimental conditions. Ordinarily, structural systematicity has been considered the domain of analogy, rather than categorisation.

In experiment 3, the influence of shared structural systematicity was remarkable. Participants preferred groupings between pots and walls, and walls and pans to pots and pans and walls alone. These findings have strong implications for categorisation and analogy research: I have noted several times the widespread acceptance in both areas of research of the two-process view of analogical / metaphorical and literal understandings, whereby 'literal' (within category) understandings are external to non-literal (analogical or metaphorical) understandings, and are therefore assumed to be computed by separate cognitive processes. The findings in this experiment seem to contradict the assumptions at the heart of this distinction.

However, in spite of this, I do not want to say at this stage that analogy *is* categorisation. It is difficult to envisage how such a central cognitive process such as categorisation could be reduced to a single process (c.f. Goldstone, 1994). Given the difficulty inherent in characterising analogical, metaphorical and categorical reasoning it seems reasonable to be as dubious of the usefulness of the kind of identity statements made by Glucksberg and Keysar, (1990), who argue that metaphorical statements should be understood as class-inclusion statements, as one should be dubious of the contrast definitions with which this account started. It is plausible - even likely - that a number of reasoning processes play a part in categorisation. The conservative interpretation of these results is to suggest that analogy is best viewed as a *sub-process* of categorisation, and not as a separate process; it may be worthwhile keeping an open mind as to whether an orange crate is an orange crate *can* be a table?

In the following chapter, I shall explore this idea further.

# Chapter 6

# Re-Examining The Role Of Semantic Constraints In Analogy

## *6.1 Introduction*

In the preceding chapters I have argued for a position that rejects any simple distinction between analogy and categorisation at the cognitive process level. To recap, the argument runs as follows: definitions of analogy (e.g. Holyoak and Thagard, 1995; Clement and Gentner, 1991) rely on making a distinction between 'straight' categorical judgements on the one hand, and analogical 'extra-categorical' judgements on the other. With no convincing account of just what constitutes a 'straight' categorical judgement, this amounts to little more than hand waving, attempting to characterise one ill-defined process by contrasting it with another ill-defined process. If the distinction is merely based upon appearances, detailed scientific scrutiny should show this. Further, it might also be that removing the distinction will allow progress made in studying analogy in isolation to inform the real question: that of explaining the processes that underpin all human conceptualisation: analogy, metaphor, and categorisation.

This chapter examines the perspective this particular view of the categorisation / analogy divide, and the results experiments that use analogical models to explore categorisation, can bring to existing theoretical approaches to analogy. The removal of neatly bounded 'concept domains' with which various aspects of the analogical process can interact produces a changed circumstance that seems likely to have repercussions for theories based on just such assumptions. In particular, I shall examine the implications of the approach developed for the semantic similarity constraints which are variously included in, or excluded from, theories and models of the analogical process.

In examining in detail the effects of the blurred analogy / categorisation perspective on 'semantics,' I shall again focus on Gentner's 'Structure Mapping' theory and Holyoak and Thagard's 'Multi-Constraint' theory. As I noted in chapter 2, these constitute the most explicit theories, whose supporting evidence and accompanying process models have been most widely disseminated and accepted. In this chapter I shall briefly restate the main elements of these theories before going on to examine the implications that the evidence , both empirical and theoretical ( presented in the intervening chapters) has for them.

---

**Solar System**

YELLOW(sun)
MASSIVE (sun)
HOT(sun)
ATTRACTS(sun,planet)
MORE_MASSIVE_THAN(sun,planet)
REVOLVES_AROUND(planet,sun)
CAUSE(ATTRACTS(sun,planet), REVOLVES_AROUND(planet,sun))
HOTTER_THAN(sun,planet)

**Hydrogen Atom**

ATTRACTS(nucleus,electron)
MORE_MASSIVE_THAN(nucleus,electron)
REVOLVES_AROUND(electron,nucleus)
CAUSE(ATTRACTS(nucleus,electron), REVOLVES_AROUND(electron,nucleus)

---

**Figure 6.1:** Predicate representations of the solar system and a hydrogen atom

## *6.2 Gentner's Structure Mapping Theory - A Recap.*

As I described in chapter 2, Gentner's Structure Mapping Theory (Gentner, 1983, etc.) proposes that the mapping and inference between two domains can be achieved by assigning correspondences between objects and attributes and then mapping predicates with identical names. The theory assumes a predicate like representation (figure 6.1), distinguishing between *objects, object-attributes* and *relations*. Object-attributes are those predicates that have one argument and describe object properties, e.g. YELLOW(sun). Relations are divided into a hierarchy of orders, with those predicates with two or more arguments which are used to describe relations between objects, for example ATTRACTS(sun, planet) forming the lowest order, and those predicates describing different levels of relationships between relations e.g. CAUSE( ATTRACTS(sun, planet),REVOLVES_AROUND(planet, sun)) forming the higher orders.

The theory comprises two parts: mapping rules, and the systematicity principle. Mapping rules state that attributes of objects are not mapped; but relations between objects are preserved. The systematicity principle requires that complex higher order relations (e.g. CAUSE in the paragraph above) are mapped preferentially, followed by relations that constitute the higher order arguments. This is intended to capture the notion that analogy conveys a system of connected knowledge, rather than an assortment of independent facts.

> "structure mapping stems in part from the observation that useful analogies, such as those used in science or education, involve rich, interconstraining systems of mappings between two domains, rather than a set of independent correspondences" (Clement and Gentner, 1991, 91-92)

In practice, effectively as a third mapping rule, structure mapping theory also insists that mappings only occur between predicates of the same name. Gentner and her colleagues (Gentner, Ratterman and Forbus, 1993; Forbus, Gentner and Law, 1995) assume the 'canonical representation of concepts' at some level of cognition, hence the question of whether, say ORBITS and REVOLVES_AROUND are matching predicates is settled outside of the theoretical parameters that structure mapping theory assumes. Gentner *et al* assume that the match between ORBITS and REVOLVES_AROUND is settled by these canonical conceptual representations, and thus can be assumed by their theory and model.[1]

---

[1] Gentner and her colleagues have latterly tended to present structure mapping theory as a more general cognitive theory of similarity (Medin, Goldstone and Gentner, 1993; Markman and Gentner, 1993; Gentner and Markman, 1997), and have even suggested - though nowhere is this suggestion fleshed out in any detail - that structural alignment might play an important part in similarity judgements in categorisation. The canonical conceptual representation assumption would appear to be a major obstacle to fleshing out these suggestions. Thus far Gentner and colleagues experiments on similarity (even those suggesting that structure mapping might play a part in categorisation - see Markman and Gentner, 1993 for examples) still follow that basic taxonomy of similarities outlined above, where differences between analogy and literal similarity judgements are determined at the *object level alone*, with structural comparitors maintaining the canonical conceptual decomposition assumption. This assumption leads to an inherent circularity when one considers structure mapping theory: if structure

As I described in chapter 2, Gentner's structure mapping theory has been implemented in a computer simulation model, SME (the Structure Mapping Engine; Falkenhainer, Forbus and Gentner, 1989).

## 6.3 A Reprise Of Holyoak And Thagard's Multi-Constraint Theory

In contrast to Gentner's theory, Holyoak and Thagard propose a multi-constraint theory (Holyoak, 1985; Holyoak and Thagard, 1989; Holyoak and Thagard, 1995) in order to attempts to explain how the large number of potential mappings between domains can be reduced to a meaningful subset for the transfer of information between domains. Holyoak and Thagard argue that this subset emerges from an attempt to balance a number of different influences upon the mapping process, the process being characterised as an attempt to simultaneously satisfy several constraints. One group of these constraints - logical compatibility, role identity, uniqueness and relational consistency - are structural in nature, and therefore compatible with the constraints posited by Gentner's theory. However, Holyoak and Thagard also argue that the constraints of *pragmatic centrality* and *semantic similarity* are involved in determining mappings during analogical processing. These non-structural constraints rely on information other than that found in the basic domain representations. In order to aid the analysis in this section, I will briefly summarise these constraints below.

### 6.3.1 Structural Constraints

The *logical compatibility* constraint ensures that mappings are only considered if they are between entities of the same 'type'. Thus, in the Solar System / Hydrogen Atom analogy the predicate MORE_MASSIVE_THAN cannot be matched with the object electron. Similarly, a mapping between the predicates HOT and REVOLVES_AROUND will not be

---

mapping plays an important part in determining conceptual judgements, then it must be a key determinant in just whether certain relational terms relate to the same 'concept' or not: whether REVOLVES_AROUND maps to ORBITS. In attempting to account for categorisation, structure mapping must assume that the phenomena it is trying to explain - conceptual mapping between match items at the relational level - is accounted for externally (by canonical concepts) in order to provide its explanation. But if structure mapping is to explain concepts, then it cannot assume that REVOLVES_AROUND maps to ORBITS canonically in order to explain how the relational concepts REVOLVES_AROUND and ORBITS match. I shall try to further illustrate this problem throughout the rest of this chapter.

considered because they take different numbers of arguments. This primarily syntactic constraint is intended to ensure that mappings between different levels of description are not attempted. For example single-argument predicates, such as HOT, tend to be purely descriptive, specifying a particular attribute of an object. Multi-argument predicates describe relationships between objects, and so can be considered to represent a higher level of description. Holyoak and Thagard argue that mappings between different levels of description are not productive and this constraint serves to eliminate any potential mappings of this kind.

The *role identity* constraint assumes that the base and target domains can be divided at a higher level of description than that at which the mapping takes place. In the use of analogy in problem solving, upon which the authors focus predominantly, this means the domains may be redescribed in terms of a start state, the problem goals, and the operators that can be used to try and achieve these goals. Role identity then limits mapping to relations and objects that appear in the same part of the domain definition. This provides a weak pragmatic influence in that elements can only be considered for mapping if they play a similar role in both domains.

Holyoak and Thagard assume that each element in the base domain will ultimately map onto only one element in the target domain, and *vice versa*. Thus there is a competition between members of the set of potential mappings between one base element and a number of possible target elements. For example, if HOTTER_THAN in the base maps onto HOTTER_THAN in the target then it cannot map onto LESS_MASSIVE_THAN in the target. Accordingly any factor which serves to increase the level of support for the former mapping will consequently act to decrease support for the latter.

The final structural constraint is *relational consistency*, which ensures the coherency and consistency of mappings between the base and target domains. If mappings between structural elements receive support, mappings between the structures themselves, and any other elements, are also supported.

### 6.3.2 Pragmatic Centrality

The importance of an element (object or relation), whether in the base or target domain, is another consideration in the mapping process. An element's importance is defined in terms of how useful the element is in satisfying the current goal (or subgoal) of the 'analogiser'. Thus any mappings involving 'useful' elements receive more support than mappings involving less useful elements. When our example analogy is used to explain the relative motion of sub-atomic particles, mappings involving YELLOW and HOTTER_THAN are going to be less favourably considered than those involving REVOLVES_AROUND, since the former are not utilised in satisfying any explanatory goals.

### 6.3.3 Semantic Similarity

Holyoak and Thagard claim that the most useful mappings in analogy are likely to come from elements which are 'semantically' similar. In the Solar System / Hydrogen Atom analogy, predicates with identical names can be regarded as more similar than those with different names. In more complex examples the method of determining relative similarity is more difficult. In their theory, Holyoak and Thagard explicitly make no claim as to any particular model of semantics; moreover, the semantic similarity constraint is regarded more as a heuristic than a firm rule, and can be applied in differing strengths at various stages of the mapping process.

Holyoak and Thagard (1989) emphasise that the logical compatibility and role identity constraints are restrictions on the building of the mapping network, and these restrictions are regarded as less important than the three principal constraints of isomorphism (the uniqueness and relational consistency constraints), semantic similarity and pragmatic centrality.

As I described in chapter 2, the theory has been implemented as a computer simulation ACME, which is a constraint satisfaction network that implements these considerations.

## 6.4 A Summary Comparison Of The Two Theories

### 6.4.1 Structure

Holyoak and Thagard (1995) argue that the similarities and differences between their theory and structure mapping theory can best be illustrated by comparing Gentner's theory with the three main constraints posited by multi-constraint theory. They claim that multi-constraint theory captures Gentner's insight regarding the importance of systematic structure (in ACME, interconnected systems will have more mutually supporting links than an isolated relation), but in a more flexible manner. As implemented in SME, Gentner's theory rigidly enforces one-to-one mappings and structural consistency: potential mappings which violate these constraints are not made. In contrast, ACME, whilst preferring one-to-one mappings (by using inhibitory links to discourage many-to-one mappings) nevertheless will allow violations.

### 6.4.2 Pragmatic Constraints

Gentner's theory (and SME) does not incorporate or recognise the influence of pragmatic - goal driven - constraints on the mapping process. According to structure mapping theory, the operation of goals is external to the actual mapping process, constraining the evaluation of mapping outcomes, rather than actual mappings (although I-SME (Forbus, Ferguson and Gentner, 1994) does incorporate some pragmatic influences in mapping). However, whilst the evidence that goals can influence *what is mapped* in analogy is clear (Spellman and Holyoak, 1992), it is less clear that goals directly influence or constrain the *mapping process*.

I noted two key criticisms of structure mapping in chapter 2: the neatness of the representations used in SME (this applies equally to ACME); and that if an analog offers up two competing possible modes of transfer with a similar level of systematicity, then the systematicity principle cannot act as a constraint in the selection of one or the other. However, as I argued in chapter 2, this latter point need not militate against structure mapping in principle: an analog which allows two equal mappings may be a poor choice of an analog, hence the systematicity principle's yielding of two equally valid mappings in such a situation may be a psychologically valid resolution of the initially poor choice of analogy. [2]

Further, it is worth recalling that some of the best evidence of the influence of goals (Spellman and Holyoak, 1992) seems to support the view that goals influence representation rather than mapping. Commenting on the results of Spellman and Holyoak's experiments examining the role of pragmatics in analogy, Holyoak and Thagard note that "[the] results suggest one way in which analogies can be used systematically to influence people's inferences - [the representation of] the source [analog] can itself be massaged to encourage a desired mapping" (Holyoak and Thagard, 1995, p 106).

---

[2] One criticism of both theories raised in Chapter 2 is that they do not adequately consider the considerable psychological evidence that the choice of representation is crucial to analogy (Hofstadter, 1995; though Hofstadter's theory is just as vulnerable to this objection); if categorisation research tends to ignore processes and overconcentrate on representation, then the opposite is true of analogy.

### 6.4.3 Semantic Constraints

With regard to the semantic similarity constraint, the main focus of this discussion, the respective positions are as follows: SME only matches predicates with identical names, thus if planets (see figure 6.1) were represented as SMALLER_THAN the sun, and electrons as LESS_MASS_THAN a nucleus, then structure mapping would not allow, and SME would not make, a mapping between the two relations. On the other hand, whilst ACME again prefers to map identical relations, weights on the network can be adjusted to capture the semantic similarity between SMALLER_THAN and LESS_MASS_THAN. As I noted earlier, Holyoak and Thagard argue that this shows a significant weakness in Gentner's theory:

> "with its emphasis on structure to the exclusion of all other constraints, SME does not simply discourage mappings between non-identical but semantically similar items; it does not even permit them." (Holyoak and Thagard, 1995, p. 258)

Holyoak and Thagard's criticism of the lack of semantic considerations in Gentner's theory carries a lot of intuitive weight. It does seem a perverse, restrictive analogical theory that rejects mappings between SMALLER_THAN and LESS_MASS_THAN in the course of an analogy mapping. However, it is not necessarily so and, from the perspective of a blurred distinction between analogy and categorisation, it might actually be that rejecting such a mapping is necessary per se, rather than necessarily perverse.

One reason for so arguing stems from the results of the investigations into the effects of systematic structure upon categorisation judgements reported in chapter 5 above. Experiments 1 and 3 addressed the question of whether analogy can be distinguished from categorisation by contrasting categorisational and analogical processes, first by presenting participants with Gentner *et al*'s analogy materials and asking them to categorise them, and then by repeating this process with novel object descriptions rather than 'analogous' stories.

Given that Gentner *et al* define the analogical mechanism in terms of structure mapping, and given the hypothesis that this process was not distinct from a basic categorisation process, I expected structure mapping to determine categorisation. Gentner *et al* assume that match items with only structural similarities (i.e. analogs) belong to different categories. It was predicted in experiment 1 that they would be categorised together: 79.5% of the groupings formed by participants in the experiment had only shared systematic structure (traditionally defined as analogy) as a common feature amongst members of the categories formed. In contrast, only 5% of groupings produced had common object descriptions as the common similarity across categories (i.e. the attribute matches often thought to be determinate of categorisation). To the 79.5% of structural congruity groupings could be added a further 8% of classifications where structural *additions* to otherwise structurally congruent representations caused them to be classed singularly.

Similar results came from experiment 3, where participants were presented with novel object descriptions in which structural similarities ran across what might be termed the objects' 'normal' categories. The results of this experiment showed considerable evidence to support the 'no distinction' hypothesis. 70% of the groupings were formed on the basis of shared structure, as opposed to just 2.5% of groupings that were classed according to shared object attributes, reflective of the 'normal' categories they embodied. Participants preferred groupings between pots and walls, and walls and pans to pots and pans and walls alone.

---

```
SMALLER_THAN
    RELATION(smaller_than)
    NO_OF_ARGUMENTS(smaller_than, 2)
    COMPARATIVE(smaller_than)
    NONREFLEXIVE(smaller_than)
    ASYMMETRIC(smaller_than)
    TRANSITIVE(smaller_than)
    CONCERNS(smaller_than, size)
LESS_MASS_THAN
    RELATION(less_mass_than)
    NO_OF_ARGUMENTS(less_mass_than, 2)
    COMPARATIVE(less_mass_than)
    NONREFLEXIVE(less_mass_than)
    ASYMMETRIC(less_mass_than)
    TRANSITIVE(less_mass_than)
    CONCERNS(less_mass_than, weight)
```

---

**Figure 6.2:** Predicate representations of SMALLER_THAN and LESS_MASS_THAN[3]

[3] Thanks to Rick Cooper for suggesting the representations in this example.

These results appear to suggest that the process of classifying two terms together (mapping SMALLER_THAN to LESS_MASS_THAN) may be no different to the process of determining the analogy between SOLAR_SYSTEM and HYDROGEN_ATOM. The process of mapping SMALLER_THAN to LESS_MASS_THAN seems to be less a sub-process of the process of determining the analogy between SOLAR_SYSTEM and HYDROGEN_ATOM and more like the same process functioning in parallel. Or, to put it another way, rather than supporting a sub-process account, the evidence reviewed suggests that mapping SMALLER_THAN to LESS_MASS_THAN involves the same processes as mapping SOLAR_SYSTEM to HYDROGEN_ATOM, except that conceptually, the two mappings occur at what one might term 'different levels of description.' Figure 6.2 presents a sketch of the kind of representations one might envisage facilitating such a parallel process, showing how SMALLER_THAN and LESS_MASS_THAN might be represented at another level of conceptual description. (I make no claims for the psychological plausibility of these representations - they are offered only to illustrate the possibility of such further expansion of the representations being mapped in an analogy.)

### *6.4.4 Prototype Schemas?*

Holyoak and Thagard claim that semantic mappings operate independently of analogical mappings. I am going to argue here that the same, or at least a significantly similar process resolves semantic differences in analogy. According to this view, the semantic mapping between SMALLER_THAN and LESS_MASS_THAN is determined by much the same process as the mapping between the atom and solar system.

One possible objection to this would be that my characterisation of categorisation is incomplete. It might be argued that analogical judgements may not be easily distinguished from classification judgements, but *categorisation* judgements can be. This claim would rely on the idea that categorisation doesn't involve that same mapping process because it makes use of generalised 'prototypical' schemas - in some way (Holyoak and Thagard advocate such a view of categories in Holland *et al*, 1984). If SMALLER_THAN and LESS_MASS_THAN share the same prototypical schema, then there would be no need to compute the similarity between SMALLER_THAN and LESS_MASS_THAN, since such similarity can be confirmed merely by reference to the prototype[4]; and it is

[4] As I noted earlier, to the extent that semantics are modelled in Gentner's account of analogy, the structure mapping view corresponds to this picture. Gentner et al (Gentner, Ratterman and Forbus, 1993; Forbus, Gentner and Law, 1995) assume some kind of canonical conceptual representation: 'SME's constraint of matching identical predicates assumes canonical conceptual representations, not lexical strings. Two concepts that are similar but not identical (such as "bestow" and "bequeath") are

this confirmatory reference-to-prototype that is modelled by the semantic links in Holyoak and Thagard's theory.

There are a number of objections to such an account, which I reviewed and discussed at length in chapter 4 above. Practically, there is the problem of providing a convincing account of what a prototype category is: no plausible or compelling theory of prototypes exist, nor does any such account seem likely (in the light of the discussion in chapter 4) to be forthcoming. Propositional schematic models of representation (some variety of which is assumed in the majority of cognitive theories) have the power to store a combination of both schemas and the exemplars from which such schemas are supposed to be constructed. As I showed in my review, what is actually stored has been the subject of much debate, and still remains, largely, an open question.

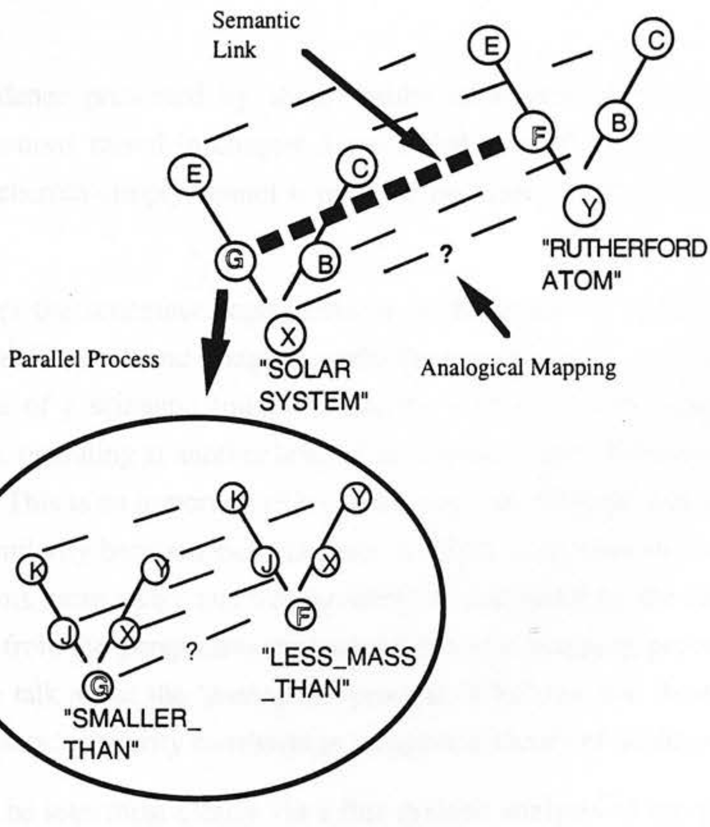



Figure 6.3: **Two views of the mapping process**

---

assumed to be decomposed into a canonical representation language so that their similarity is expressed as a partial identity (... "give")' (Gentner Ratterman and Forbus, 1993); conceptual matters are thus explicitly extrinsic to both structure mapping's theory and models.

Experiment 2 was designed to see whether analogical theory could shed any light on the nature of participants' stored representations. It yielded little evidence to support the hypothesis that participants had abstracted and stored schemas from the groups they had classified, despite the fact that it appeared to be a shared structural schema that was the basis of participants' original classification decisions (experiments 1 and 3, above). Instead, when Gentner, Ratterman and Forbus's, 1993, analysis was applied to participants' behaviour, it appears that being presented with a class name in no particular context caused the participants to randomly recall one of the exemplars associated with that name, and then use that exemplar as the stimulus for recalling other class members. Thus as well as finding no evidence to support stored prototypes, the results of experiment 2 provide some tentative evidence that it is possible to use class-names when only exemplars (instances) of categories are stored (which strengthens the argument for not strongly distinguishing analogy and categorisation).

Whilst the evidence presented by these results is somewhat tentative, when the theoretical objections raised in chapter 5 are added as well, it seems reasonable to conclude that schemas simply cannot supply the necessary theoretical justification for semantic links.

If one considers the schematic representation of the mapping process presented in figure 6.3 above, Holyoak and Thagard model the mapping of F onto G in analogy X -> Y by means of a semantic link, whereas the evidence points towards a similar parallel process, operating at another level of conceptual description for mapping F -> G and X -> Y. This is an important distinction: whereas Holyoak and Thagard model the semantic similarity between SMALLER_THAN and LESS_MASS_THAN as a *sub-process* of analogy, it seems more likely that this similarity is computed by the *same process* in parallel. Since from the perspective presented here, this mapping process is what we mean when we talk about the 'analogical' process, it follows that there simply is no place for a semantic similarity constraint in a cognitive theory of analogy.

Again, this can be seen most clearly via a fine-grained analysis of the predicates as in figure 6.2. If explaining how semantics are reconciled or mapped involves a recursive or parallel call to the same process, then it makes no sense for semantic mapping to be part of that process. In explanatory terms, such an account is circular: it proposes that semantics are mapped by a process which utilises semantic links (but these represent semantic mappings in the first place).

## 6.5 Two Views Of Semantic Similarity

According to Holyoak and Thagard, semantic mappings operate independently of analogical mappings. According to the analysis presented here, the same, or at least a significantly similar process may be operating. According to this view, the semantic mapping between SMALLER_THAN and LESS_MASS_THAN is determined by much the same process as the mapping between the atom and solar system. Holyoak and Thagard offer no theory or model of how semantic links are processed. On the other hand, it could be argued that although I have advanced a theoretical analysis above, the process model I advocate still amounts to little more than hand-waving. The following pair of studies were designed to add flesh and empirical support to the speculative model of semantic processing in analogy presented in this chapter.

## 6.6 Semantic Mapping As A Parallel Process

Experiment 4 was designed to make concrete this proposal, and explore it empirically. As cast by Holyoak and Thagard, the question of semantics revolves around supplying an account of what happens when two 'semantically similar' terms - SMALLER_THAN and LESS_MASS_THAN - are encountered during the mapping process. In ordinary usage, the representations of human category information involved in these processes are implicit; people know what SMALLER_THAN and LESS_MASS_THAN mean, and they reconcile (or map between) the two terms accordingly. But the exact nature of *what* they know, and *how* such knowledge is represented appears to be inaccessible at the level of detail required to specify and model the underlying process involved in their 'semantic reconciliation' of the two terms. In the following study, participants were asked to make inferences with the aid of two targets (figure 6.4). However, in the base and each of the targets, the term that was crucial to determining the representation of higher order structure in the scenarios was a novel, artificial term. By supplying 'definitions' for that term, we hoped to be able to control the representations participants used for semantically reconciling particular terms during their analogising. The prediction tested was that participants would use the same process to match semantic items in their representations as they would in ultimately determining their analogies (as described above, and illustrated in figures 6.2 and 6.3).

## SCENARIOS

### BASE - The Guralaga

***can be found in Australia***
***lives in Rainforests***
***only eats gau-gau berries***
has a cronomus lucundus
the *cronomus lucundus* enables the Guralaga to eat gau gau berries.

### TARGET 1 - The Mongret

***can be found in Australia***
***lives in Rainforests***
***only eats gau-gau berries***
has a probus razoris
the probus razoris enables the Mongret to eat the gau gau berries.
Thanks to the way they eat, Mongrets live to a ripe old age and rarely suffer from cancer

### TARGET 2 - The Crany Dog

can be found in Papua new Guinea
lives in the grassy backlands
eats vegetation
has a remulum grandoso
because of the *remulum grandoso* the Crany Dog can eat vegetation.
*Crany Dogs are particularly prone to cancer, which originates in their digestive system.*

**Figure 6.4:** A base and two targets. The surface similarities between the base and the SMT are highlighted. The base and the SST share few surface similarities.

**"DICTIONARY ENTRIES"**

**BASE DICTIONARY ENTRY**

**Cronomus lucundus**

***are unique to certain types of bird***
***are important to berry eaters***
***is a long*** spleen-like organ
*keeping berries in the cronomus lucundus allows the berries to slowly ferment, allowing the goodness inside the bitter skins to be released*

**SMT DICTIONARY ENTRY**

**Probus razoris**

***are unique to certain types of bird***
***are important to berry eaters***
***is a long*** plier-like bill
crushing berries in the probus razoris allows the goodness inside their bitter skins to be released without the skins having to be swallowed

**SST DICTIONARY ENTRY**

**Remulum grandoso**

is unique to certain types of dog
are important to dogs which eat a wide range of vegetation
is a short intestine-like organ
*keeping vegetation in the remulum grandoso allows it to slowly ferment, allowing the goodness inside the outer skins to be released*

**Figure 6.5**: Dictionary entries for the Base and two Targets in figure ??. Surface similarities between the Base and SMT are in bold italic print; the structural match between the Base and the SST is in normal italic

## 6.7 Experiment 4

### 6.7.1 Participants

The participants were 170 volunteers, a mixture of postgraduate and undergraduate students from the Department of Artificial Intelligence, Centre for Cognitive Science, Department of Psychology and the Faculty of Music at the University of Edinburgh.

### 6.7.2 Materials And Design

The materials comprised 5 groups of specially constructed scenarios (figure 6.4) with corresponding sets of novel dictionary entries (figure 6.5) and the 2 candidate inferences for each group (figure 6.6). Each scenario group was further sub-divided into two versions of the scenario sets, and two versions of the dictionary entry sets, so that each scenario / dictionary sub-set supported one of the two different candidate inferences. As in the earlier experiments, Gentner, Ratterman and Forbus's (1993) taxonomy of similarity relationships was used to classify the similarities supported by the dictionary entries and within the scenario sets:

- *Literal similarity* matches include both common relational structure and common object descriptions;
- *Surface matches*: based upon common object descriptions, plus some first order relations;
- *Structural similarity*, matches based upon a common system of relations.

Thus the relations between the various scenarios in a given scenario group can be summarised as follows (see also figure 6.7):

*Group type A:* the base and one target scenario (the SST, or structurally supported target) shared only structural matches; mappings between the SST's dictionary entry and the base dictionary entry also shared only structural matches. There was a structural correspondence between the base structure supported by the base dictionary entry and the SST's dictionary entry which supported candidate inference A. Matches between the base and the other target scenario (the SMT, or surface match supported target) include surface matches and mappings between the SMT's dictionary entry and the base dictionary entry also shared common object descriptions. There was a structural correspondence between the base structure supported by the base dictionary entry and the SMT's dictionary entry which supported candidate inference B.

*Group type B:* matches between the base and the SST were structural; mappings between that SST dictionary entry and the base dictionary entry also included only structural matches - there was a structural correspondence between the base structure supported by the base dictionary entry and that supported by the SST's dictionary entry which supported candidate inference B. Matches between the base and the SMT were based on surface matches; mappings between that SST dictionary entry and the base dictionary entry also included only surface matches - there was a structural correspondence between the base structure supported by the base dictionary entry and that supported by the SMT's dictionary entry which supported candidate inference A.

---

**INFERENCES**

***A.*** *Guralaga live to a ripe old age and rarely suffer from cancer .*

***B.*** *Guralaga are particularly prone to cancer.*

---

**Figure 6.6:** The target inferences for the stimulus group shown of the following pages. In a type A set, structural commonalities would support the A inference surface similarities would support the B inference. In a type B set, structural commonalities would support the A inference surface similarities would support the A inference.

To try and simplify the above - in each group of stimuli, the base and one target scenario, and their corresponding dictionary entries, shared surface features, and a higher order structural correspondence that corresponded with one candidate inference, whilst the base and the other target scenario, and their corresponding dictionary entries, shared structural correspondences, and a higher order structural correspondence that corresponded with the alternative candidate inference.

Each stimulus set was divided into two subsets: in one, structural features in the targets and their novel term dictionary entries supported one set of inferences (Type A), whilst in the second sub-set, the same kind of matches supported the contrasting inference (Type B), so that biases towards a given inference could be eliminated (see also figure 6.7).

In keeping with the analysis presented above, we expected that participants would use analogy to reconcile semantic terms in order to perform analogical mappings between the scenarios and generate support for one candidate inference. We predicted that in order to be able to carry out the top level analogy, participants would carry out another analogy in parallel - mapping structures only in the dictionary entries - reconciling semantic terms in a way that supported the top-level 'analogical' structure mapping over the top-level surface mapping, and favour the inference that corresponded to the structurally similar scenario over the scenario that shared only surface features.

In addition to the basic stimuli, 3 sets of control stimuli were also created:

1 In the main control, the dictionary entries were eliminated, and participants were given only the Base and the two targets. In this control, in the absence of any structural support from the dictionary entries for the SST inference, we expected participants to use the surface commonalities between the base and the SMT to determine their inference choice (i.e the prediction was that when subjects were asked to make an inference in a situation where neither of the target inferences benefitted from any structural bias, partcipants would prefer the inference which was additionally supported at the object level to the inference that received no such support; consistent with the findings of previous studies, such as GentnerRatterman and Forbus, 1993, we expected weak similarity to provide more support than no similarity).

2 In the second control, participants were given materials in which the novel terms were removed, and the structural information in the dictionary entries was added to the base and target - in effect creating "normal" analogy materials. In this control, again consistent with previous findings such as Gentner, Ratterman and Forbus (1993), we expected the structural commonalities between the base and the SST to determine the choice of inference, overriding the surface commonalities between the base and the SMT (see figure 6.8).

3 In the final control set the dictionary entries were altered so that surface and structural commonalities all supported the same mapping (the LST, or literally similar target). In this final control, both structural and surface commonalities between the base and the LST, and their dictionary entries were aligned in support of one. Since structure was predicted to be the key factor in deciding inferences (in line with the findings of previous studies), I did not expect the results from this control to differ significantly from the main experimental task.

In each of the controls, the inference supported by the various similarities was again alternated to remove any influence of inherent biases towards given inferences.

| | **Surface Match Target (SMT)** | **Structurally Similar Target (SST)** |
|---|---|---|
| **Analogy Level** | Shares **surface** features with base<br><br>***Structral*** overlap with base determined by dictionary mapping | Doesn't share surface features with base<br><br>***Structral*** overlap with base determined by dictionary mapping |
| **Dictionary Level** | Shares **surface** features with base<br><br>Doesn't share structure with base | Doesn't share surface features with base<br><br>Shares **structure** with base |
| **Inference** | In ***type A*** sets, the ***B*** inference is only supported by **surface** matches between base and SMT the dictionary entries<br><br>In ***type B*** sets, the ***A*** inference is only supported by **surface** matches between base and SMT the dictionary entries | In ***type A*** sets, the ***A*** inference is only supported by **structural** matches between base and SST the dictionary entries<br><br>In ***type B*** sets, the ***B*** inference is only supported by **structural** matches between base and SST the dictionary entries |

**Figure 6.7a:** The relationships between the base, targets, dictionary entries and inferences in the main stimulus groups.

**Chateau Bogusse:**

**is a vineyard.**
**is in the southern French district of Pretence.**
**has sandy soils, with a lot of surface pebbles**
**has a warm microclimate which enables grapes to be produced.**
the particular microclimate results in ripe grapes.
*the ripeness causes the sugar level in the grapes to rise.*
*this makes the walls of the grapes weaken and collapse.*

**Domaine Fraudulent:**

grows plums.
has clay soils in which wildflowers grow
is in the western Departement of Maidoop.
its warm microclimate causes melons to grow
the particular microclimate yields extremely ripe plums.
*the extreme ripeness causes the plums to become very sweet*
*this super-sweetness makes the plums soft and squashy*
*Because of their squashiness Domaine Fraudulent's plums are held in low esteem, and sell poorly.*

**Mas de la Fiction:**

**grows grapes.**
**is in the southern Departement of Whaupper.**
**has sandy soils, with a lot of surface pebbles**
**its fine microclimate causes grapes to grow.**
**the particular microclimate results in ripe grapes.**
the ripeness causes some of the moisture in the grapes to evaporate
this evaporation leads to extremely concentrated flavours
Because of their concentrated flavours Mas de la Fiction's grapes are prized and sell for high prices.

**Inferences**

A. Chateau Bogusse's grapes are highly prized and sell for high prices.

B. Chateau Bogusse's grapes are held in low esteem, and sell poorly.

**Figure 6.8:** Control set 2, in which the structural information in the dictionary entries has been included in the base and targets to create an "ordinary" analogical problem. Surface similarities are illustrated in bold; structural similarities are italicised.

### 6.7.3 Procedure

Participants were presented with 2 x 6-page questionnaires, each of which contained one scenario set, with its dictionary and candidate inferences, a diversionary task and a scenario set and pair of candidate inferences without a dictionary (the main control). The order in which the sets were presented ('with-dictionary' versus 'without-dictionary' control), was randomised, as was presentation order of the targets within the sets. A second, smaller group of participants were given the other two controls in similar fashion.

Participants were asked to infer one candidate inference, and give a confidence rating (1=not at all confident; 5=very confident). They were told that the dictionary entries might be useful to them, but told explicitly that the use of them was left up to the individual participant's discretion.

### 6.7.4 Results

Consistent with the initial hypothesis, in the main control condition where no dictionary entries were provided, the inference which received common surface-feature support was favoured by 67% of participants, with only 33% preferring the inference that was supported by structure commonalities, $\chi^2(1, N=140) = 17.1$, $p<.001$.

However, when definitions, which offered the possibility of structural mappings, were provided, participants reversed their preferred inference for a given base / targets set. Again consistent with the initial hypothesis, in this condition, if participants had preferred the A inference in the first control, when provided with scenario sets where structural commonalities in the dictionary supported the B inference, then participants now chose the B inference. Overall the inferences which received structural support were favoured by 71.2% of participants, with only 28.8% preferring the inference that was supported by surface commonalities alone, $\chi^2(1, N=125) = 20.748$, $p<.001$.

Also consistent with the initial hypothesis, in the control condition with no novel terms, where structure in the dictionary entries was included in the base and targets, inferences which received common surface-feature support were favoured by only 27.0% of participants, with 73.0% preferring the inference that was supported by structural commonalities, $\chi^2(1, N=26) = 3.869$, $p<.05$.

There was no deviation from this pattern in the final control condition, where the dictionary entries were altered so that surface and structural commonalities all supported the same base - target (the LST) mapping, the inference supported by the LST was favoured by 75.0% of participants, $\chi^2$ (1, N=28) = 5.17, $p<.05$

Analysis of participants' confidence scores in the main control show significantly greater confidence for inferences based on surface commonalities when no structure was present, $t=8.72$, $p<0.001$. However, this trend was reversed in the other controls and the main experiment - given the choice, participants seem to prefer structurally supported inferences. In the second control condition (analogies) inferences based upon structural commonalities received a significantly higher confidence rating than those based on surface features, $t=3.982$, $p<0.001$. Similarly, when definitions were provided, inferences based upon structural support received a significantly higher confidence rating than those based only on surface commonalities, $t=2.9$, $p<.005$. This trend was repeated in the third control, though mean differences were not significant, $t=1.02$, $p=0.33$.

### 6.7.5 Discussion

Experiment 4 has shown, consistent with the claim that analogical and conceptual mappings cannot be distinguished at a cognitive process level, that participants can use the same process that they used to process and make analogical inferences to process and reconcile semantic discrepancies they encountered in the representations of base and target analogs.

Participants made inferences with the aid of two targets. By controlling the structure of the information representing the "semantics" of the term that was in turn crucial to the determination of the representation of higher order structure in the base and each of the targets, we were able to control the representations participants used for semantically reconciling particular terms during their analogising. The 'no distinction' prediction - that participants would use the same mapping process to match semantic items in their representations as they would in ultimately determining their analogies - is strongly supported by the results of this experiment.

## 6.8 "On-Line" Parallel Processing

Two very reasonable objections might be made to the results of experiment 4:

1. Firstly, the "dictionary entries" in the main task were artificial. Moreover, I have claimed in chapter 4 that definitions are an inadequate basis for conceptual semantics.

Since the "dictionary entries" appear to be little more than definitions, I appear to be wanting to have my cake and eat it. I have argued that concepts cannot be simple definitions, so how can definitions serve as the basis for exploring conceptual reconciliations in analogical mapping?

2. Notwithstanding the first objection, a second obvious objection to the findings of experiment 4 is that participants were presented with the tasks on paper, and had unlimited time in which to solve the inferencing problems and reconcile and map any "semantics" in the various base and target specifications. It might be said in objection that since structure mapping is a computationally expensive process - especially in comparison to mapping surface features - experiment 4 has little relevance to the on-line demand characteristics of analogical processing "in the wild". Since participants in experiment 4 had unlimited time, and external representations of the problems, their behaviour is no predictor of the kind of processes used in making analogical mappings "in the head", where working memory limits may place severe restrictions on processing capacity.

In order to meet the objections one might reasonably raise with respect to the findings of experiment 4, it would appear that what is needed is two concepts that can both support contrary systems of inferential support from a single lexical term (to mimic the reversal of inference patterns prompted by the materials in experiment 4) *and* support interpretations of one another analogically (to capture the structural determination of mappings between dictionary items in experiment 4). This might, at first glance, seem a difficult - perhaps impossible - requirement to meet. Except perhaps for the rarefied world of Borges stories and cognitive psychology experiments, one might think that *no* concept could meet the need for systematic contradiction, let alone the *two* concepts that will be needed to meet the demand for analogical priming. However, despite perhaps one's natural intuitions to the contrary, there are two - at least - common, everyday conceptual domains that do appear to meet these requirements: *space* and *time*.

### 6.8.1 Space And Time - A "Domain Of Semantics"

There is a great deal of overlap in the lexical terms we use in talking about space and time. We can talk of looking *forward* to the author getting to the point; we might look *ahead*, to that *far off* time when the author does get to the point, and the waffling pre-amble is *behind* us, safely in the past. A number of researchers have noted these correspondences between the words we use in talking about space and time, and have argued that the connection between the two is more a case of systematic connection rather than a mere linguistic coincidence (Fillmore, 1971; Clark, 1973; Traugott, 1978; Lakoff and Johnson, 1980; Boroditsky; 1998). The claim is that the presence of these and the presence of other systematic correspondences between 'domains'[5] in language is indicative of the existence of *conceptual metaphors*; occurrences where language from one 'domain' is used in other 'domains' (Fillmore, 1971; Lakoff & Johnson, 1980, Lakoff, 1987a; Boroditsky; 1998). Such metaphors are supposed to reveal a particular source-to-target mapping: the use of language from the domain of 'travel' in the domain of 'love' is taken to be indicative of some systematic analogical mapping between the concept 'travel' and 'love' ('love is a journey', etc.; see Lakoff 1987a; Lakoff and Johnson, 1999).[6]

One area in which these linguistic patterns have been strongly delineated is in the event sequencing aspect of time, and its relation to similar spatial concepts. Traugott (1978) defines temporal sequencing as 'the system whereby events and situations $E_1$, $E_2$, ... $E_n$ are ordered with respect to each other' (Traugott, 1978, p 379), i.e. "one's best days ($E_1$) may be *ahead* of one ($E_2$)"; "port and cigars ($E_1$) might *follow* dinner ($E_2$)".

As McTaggart (1908; 1927) argues, the concept of time has a certain inherent structure:

> "Positions in time, as time appears to us *primâ facie*, are distinguished in two ways. Each position is Earlier than some and Later than some positions. To constitute such a series there is required a transitive asymmetrical relation, and a collection of terms that, of any two of them, either the first is in this relation to the second, or the second is in this relation to the first. We may take here either the relation 'earlier than' or

[5] I shall continue my practice here of treating the rather woolly 'domain' as being equivalent to a 'concept', or set of 'concepts'.

[6] Lakoff and Johnson (1980; 1999) argue that the human conceptual system is structured around a small set of experiential concepts; concepts that emerge directly out of experience and which are defined "in their own terms." These fundamental concepts include basic spatial relations (up/down, front/back), a set of "physical ontological" concepts (entity, container), and a set of basic experiences or actions (eating, moving). This view, which seeks to ground all other concepts in these basic experiential concepts, proposes that concepts that do not emerge directly out of physical experience must be metaphoric in nature, and that these "abstract" concepts are represented via on-line mappings to the fundamental experiential concepts (Lakoff & Johnson, 1980, 1999; Lakoff, 1987).

> the relation 'later than', both of which, of course, are transitive and asymmetrical. If we take the first, then the terms have to be such that, of any two of them, either the first is earlier than the second, or the second is earlier than the first."
> McTaggart, (1927 , pp 24).

> "each position is either Past, Present, or Future. The distinctions of this former class are permanent., while those of the latter are not. If *M* is ever earlier than *N*, it is always earlier. But an event, which is now present, was future, and will be past."
> McTaggart, (1908 , p 458).

Later researchers have adopted this basic framework, generally conceiving of time as a one-dimensional, directionally oriented phenomenon, and have thus argued as a result that the spatial terms 'imported' from the domain of space to the domain of time tend to be one-dimensional (*front / back*; *up / down*) rather then multi-dimensional (*narrow / wide*) and determinably orderable (*front / back*) rather than symmetric (*left / right*) (Boroditsky, 1998). Boroditsky notes that crosslinguistically this pattern is relatively stable, with spatial terms relating to front / back relations being the ones most widely borrowed to talk about time across languages (Clark, 1973; Traugott, 1978).

Two 'metaphoric systems' have been identified for talking about event sequencing in English (Fillmore, 1971; Clark, 1973; Traugott, 1978; Lakoff and Johnson, 1980; Boroditsky, 1998), the *ego-moving* metaphor system, and the *time-moving* metaphor system.[7]

#### *6.8.1.1 The* Ego-Moving *Metaphor*

The first of these metaphoric systems is the *ego-moving* metaphor, where the *ego* (an observers' viewpoint or context) is conceived of as progressing along a 'time-line' towards the future, as in "we are *approaching* the millennium" (see also figure 6.9).

#### *6.8.1.2 The* Time-Moving *Metaphor*

The second metaphoric system in which spatial terms are used to talk about time is the *time-moving* metaphor, in this metaphor, 'a time-line is conceived of as a river or

[7] McTaggart (1908) also observes that two 'metaphoric systems' have been identified for talking about event sequencing.

conveyor belt on which events are moving from the future to the past' (Boroditsky, 1998, p 309), as in "the millennium is *coming*" (see also figure 6.10).

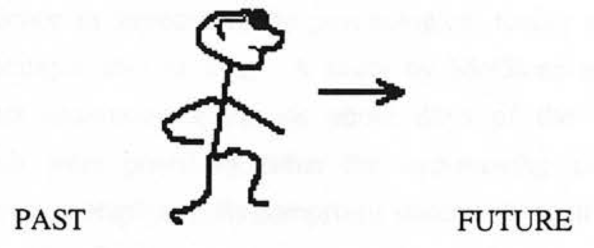



**Figure 6.9** Schematic representation of the ego-moving schema used in time sequencing (from Boroditsky, 1998).[8]

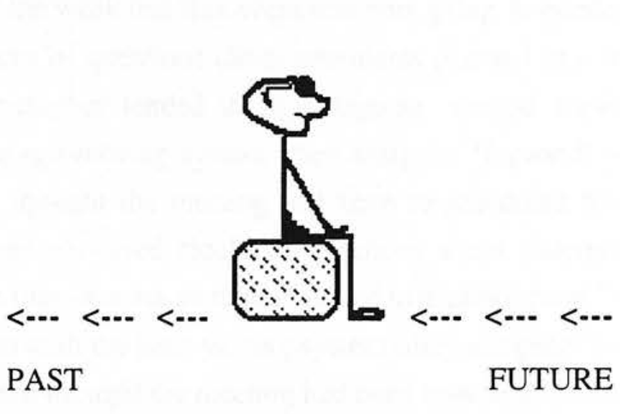



**Figure 6.10** Schematic representation of the time-moving schema used in time sequencing (from Boroditsky, 1998).

### *6.8.1.3 Time- And Ego-Moving Systems*

According to this view, these two metaphoric systems lead to two different assignments of *front* and *back* relative to a time-line (Fillmore, 1971; Clark, 1973;

[8] I am grateful to Lera Boroditsky for allowing me the use of these and later diagrams.

Traugott, 1978; Lakoff and Johnson, 1980; Boroditsky, 1998). In the *ego-moving* metaphor *front* is assigned to the **future**, or later event; e.g. "success lay *behind* them; failure lay *ahead* of them."). On the other hand, in the *time-moving* metaphor system, *front* is assigned to the **past**, or earlier event; '"I will see you *before* 4 o'clock." or "The reception *after* the talk.",' (Boroditsky, 1998, p309).[9]

There is some evidence in support of the psychological reality of there being two systems for sequencing events in time. A study by McGlone and Harding (1998) involved participants answering questions about days of the week - relative to Wednesday - which were posed in either the *ego-moving* or the *time-moving* metaphor. *Ego-moving* metaphor trials comprised statements such as "We passed the dead-line two days ago", whilst *time-moving* metaphor trials involved statements such as "The dead-line was passed two days ago"; in each case, participants read the statements and were then asked to indicate the day of the week that a given event had occurred or was going to occur. At the end of each block of such statements, participants read an ambiguous statement, such as "The reception scheduled for next Wednesday has been moved forward two days"[10] and then were again asked to indicate the day of the week that this event was now going to occur. Participants who had answered blocks of questions about statements phrased in a way consistent with the *ego-moving* metaphor tended to disambiguate "moved forward" in a manner consistent with the *ego-moving* system (they assigned 'forward' - the front - to the future, and hence thought the meeting had been re-scheduled for Friday), whereas participants who had answered blocks of questions about statements phrased a way consistent with the *time-moving* metaphor tended to disambiguate "moved forward" in a manner consistent with the *time-moving* system (they assigned 'forward' - the front - to the past, and hence thought the meeting had been re-scheduled for Monday).

---

[9] It is worth noting that the assumption here is that *before* and *after* are spatial terms - i.e. they literally denote spatial concepts - since they are 'derived from expressions relating to spatial orientation' (Fillmore, 1971, p46; see also Traugott, 1978). Boroditsky (personal communication, 1999) calls them 'spatio-temporal' terms. I shall argue later that these different assumptions have differing implications, dependent upon the theoretical baggage that accompanies them.

[10] All trials were conducted on a Wednesday.

### 6.8.2 If There Are Two Systems Of Space And Time, What Kind Of Systems Are They?

McGlone and Harding's study provides some evidence that in talking about *time*, at least, the requirement for a concept that yields contrary systems of inferential support from a single lexical term (to mimic the reversal of inference patterns prompted by the materials in experiment 4) can be met.

However, as Boroditsky (1998) notes, despite the apparent systematicity in the connection between spatial and temporal modes of expression, it does not necessarily follow from this that 'structured conceptual schemas are necessary to process metaphoric expressions about time' (Boroditsky, 1998, p 309). Whilst McGlone and Harding present evidence to support two systems of understanding time (thereby meeting the first part of the requirement for 'in the wild' concepts that could allow a replication of experiment 4), they do not provide any evidence that time can be understood as a metaphor from space (which might allow the second requirement on real world concepts to be met: that the concepts that meet the need for systematic contradiction could also meet the demand for analogical priming).

A number of explanations that are consistent the results of McGlone and Harding's study could be plausibly put forward. With the idea of resolving speculations as to the kind of explanations one might advance for this, Murphy (1996) offers up three possible frameworks through which findings such as McGlone and Harding's might be interpreted: the strong metaphoric representation view; the weak metaphoric representation view; and the structural similarity view. Each embodies an alternative explanation for the existence of systematic conceptual metaphors.

#### 6.8.2.1 Strong Metaphoric Representation (Strong MR)

According to strong MR, concepts are not understood via their own representations, but instead by metaphoric reference to another domain (i.e. represented only by links to concepts that *are* actually represented as entities). According to this view, *space* concepts are represented independently, and *time* concepts are represented by links to *space* concepts. A consequence of this view would be that temporal reasoning would involve some kind of on-line alignment between a representation of a temporal construct (say in the interpretation of a temporal phrase) and an appropriate spatial concept. Thus this view would appear provide theoretical support for structural (analogical) priming between the representations of real world concepts associated with real world lexical items, as opposed to the invented dictionary definitions in experiment 4.[11]

#### 6.8.2.2 Weak Metaphoric Representation (Weak MR) 1

Weak MR accepts that abstract concepts are represented, but claims that the existence of systematic verbal metaphors in our culture has influenced the structure of our concepts so that they are consistent with the metaphor:

> "metaphors have some influence on the representation of the topic [target] concepts, but these concepts nonetheless have their own separate representations. That is, the metaphor may influence the structure of the topic concept, but the representation itself is not metaphoric" (Murphy, 1996, p 178)

Although Murphy only articulates one interpretation of this position, it is actually consistent with two rather different interpretations, which I shall call Weak MR1 and Weak MR2. Weak MR1 corresponds to Murphy's Weak MR; I will discuss Weak MR2 after presenting the third of Murphy's alternative frameworks, the structural similarity view (SS).

*Weak MR1*: Weak MR1, which is consistent with the position put forward by Murphy, holds that no aspect of conceptual metaphors is metaphoric: the *ego-moving* and *time-moving* systems used for sequencing events may have originally been mapped metaphorically from space, but they have now become conventionalised such that *time* concepts are no longer connected to *space* concepts via an on-line mapping. On this view, the similarities apparent between space and time are etymological relics, indicative of on-line processing in our forebears rather than any current on-line connectivity between the two conceptual schemes. Thus, according to Weak MR1, space and time will not meet the analogical priming requirement for a realistic replication of experiment 4, because no on-line structure mapping will take place between them. According to this view, words like *ahead* and *forward* are merely polysemous, in that they have two concepts associated with them (roughly 'spatial front' and 'temporal front').

[11] This view would also support the proposition that 'abstract domains' such as time are structured and understood according to metaphorical mappings to experiential domains such as space (Lakoff and Johnson, 1980; Lakoff, 1987; Lakoff and Johnson, 1999).

### 6.8.2.3 The Structural Similarity View (SS)

SS holds that all concepts are directly represented, and that furthermore, metaphoric ways of talking do not reflect a general influence of such metaphors on the representation of abstract concepts. Instead, Murphy claims, metaphors arise out of similarities between pre-existing structures. Metaphors like

*argument is war*

arise out of structural similarities between the concepts *war* and *argument*. That is, Murphy claims that metaphorical ways of speaking arise out of the recognition of structural similarities between pre-existing domains. According to this view, space and time are represented separately, and the fact that people use similar terms to talk about the two is simply reflective of this coincidental underlying structural similarity:

> 'This [coincidental] structural similarity permits people to construct understandable verbal metaphors. Those that are the most interesting or revealing "stick" and may become conventional ways of talking. Those that are unrevealing or poor correspondences do not stick and so do not become conventional"
> (Murphy, 1996, p 179)

Like Weak MR1, this view sees the use of the same lexical terms in talking about space and time as resulting from historical structural alignments between space and time, but on this view, rather than these alignments having played some causal role in the shaping of one of the concepts, in this case the structures precede such talk, rather than reflect its influence. Like Weak MR1, SS is incompatible with the use of space and time to meet the analogical priming requirement in a realistic replication of experiment 4. Again, according to this view, no on-line structure mapping takes place between them. (According to this view, words like *ahead* and *forward* are strongly polysemous, in that they have two concepts associated with them -roughly 'spatial front' and 'temporal front'- purely as a result of chance structural similarities between those concepts.)

### 6.8.2.4 Weak Metaphoric Representation (Weak MR) 2

*Weak MR2*: One rather bemusing aspect of the framework put forward by Murphy is it seems to propose a very simplistic, componential view of concepts. Conventionalised metaphors are either concepts which are mapped on-line to neatly structured representations of other concepts (Strong MR), or they are neatly separate concepts. which are neatly representative of some long finished and fossilised on-line mappings that influenced their structure (Weak MR1); or else they are indicative of some long finished and fossilised on-line mappings between two already neatly represented concepts which resulted in the use of similar terms "sticking" (SS). None of this sits well with the view of concepts presented in chapter 3, or with the empirical evidence reviewed in chapter 4. All of the views put forward by Murphy presuppose some form of neatly structured, static conceptual representation. In contrast to the picture painted by that evidence, strong MR presupposes neat, systematically represented experiential concepts which can provide systematic structure for abstract concepts; Weak MR1 and SS propose that any on-line aspects to conventionalised metaphors are firmly relegated to history. Given that the evidence reviewed in chapter 4 seems to indicate quite strongly that such neat conceptual representations might not be psychologically likely (if such widespread conceptual neatness existed, one might perhaps expect to see some evidence of it), Weak MR2 posits that that some parts of conceptual metaphors are metaphoric in the manner of Strong MR, whilst other parts may be reflective of prior structural mappings that have become conventionalised, whether in the manner of Weak MR1, or SS, or both (Weak MR2 could even allow for the idea that - perhaps through some kind of symbiosis - the structure of both concepts could be affected by the process of mapping). According to this view, some aspects of the *ego-moving* and *time-moving* systems used for sequencing events may have originally been mapped metaphorically from space but they have now become conventionalised such that those particular *time* concepts are no longer connected to *space* concepts via an on-line mapping, whilst other aspects of *ego-moving* and *time-moving* systems used for sequencing events may still be mapped metaphorically from space on-line.

Accordingly, Weak MR2 also provides theoretical support for structural (analogical) priming between the representations of real world concepts associated with real world lexical items, as opposed to the invented dictionary definitions in experiment 4, but such support would not be universal. Weak MR2 would be consistent with the idea that temporal reasoning *could* involve some kind of on-line alignment between a representation of a temporal construct (say in the interpretation of a temporal phrase) and an appropriate spatial concept, but on this view, whilst such mappings would be sufficient for temporal understanding, they would not be necessary to it.

### 6.8.3 Evidence For On-Line Mapping Between Space And Time

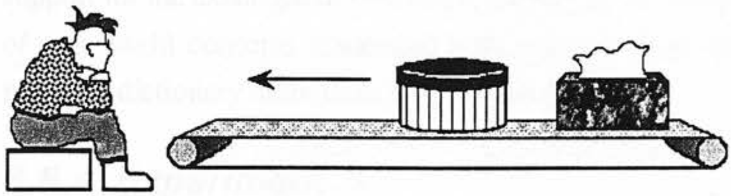
As the foregoing speculation may indicate, whilst much has been made of these linguistic patterns, little evidence has been advanced to support their psychological viability as a plausible representation scheme for conceptual knowledge (and all of them are consistent with the findings of McGlone and Harding discussed above). However, in a recent study Boroditsky (1998) has attempted to provide some empirical support for the metaphoric representation hypothesis by focusing on the event sequencing aspect of time, and its relation to similar spatial concepts, as described above.

Extending the paradigm used by McGlone and Harding, Boroditsky defines an explicit analogy between two schemas for organising space and time. On this analogy, *ego-moving* schemas are defined - for both space and time - in respect to an observer's direction of motion. The 'front' is assigned as the furthest forward point in the observer's direction of motion, thus in time, 'front' is assigned to the future, and in space, if objects are conceived of in linear fashion along a path, then 'front' is assigned to the objects at furthest forward - relative to the observer's direction of motion - along the path. For *time-* and *object-moving* schemas, front is set to the furthest forward point in the direction of the movement of time or objects. Since time is usually conceived of as moving from future to past, 'front' is assigned to past, or earlier events. By analogy, in space, if two objects are moving (whether they have intrinsic 'fronts' or not),[12] then front is assigned to the leading part of the leading object (see figure 6.11)

Boroditsky used a similar target and task to the one used in McGlone and Harding (1998)'s study ("Next Wednesday's meeting has been moved forward two days"; what day is it now on?") and these explicit analogies between space and time to demonstrate that spatial schemas could indeed prime temporal reasoning on-line. Each target question was preceded by several prime questions that used either the ego-moving schema or the object/time-moving schema (see figures 6.11 and 6.12 for example primes). Participants answered a series of TRUE/FALSE questions about the primes, and then the target question. Like McGlone and Harding, Boroditsky found that participants who had answered blocks of questions about temporal primes for the *ego-moving* metaphor tended to disambiguate "moved forward" in a manner

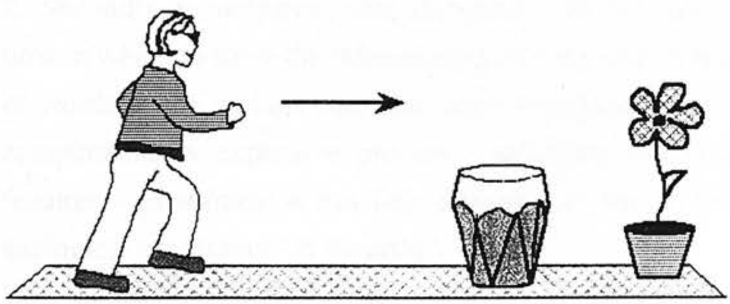
[12] Thus, when a car reverses, for instance, the back of the car could be said to be in front.

consistent with the *ego-moving* system (they assigned 'forward' - the front - to the future, and hence thought the meeting had been re-scheduled for Friday), and participants who had answered blocks of questions about temporal primes for the *time-moving* metaphor tended to disambiguate "moved forward" in a manner consistent with the *time-moving* system (they assigned 'forward' - the front - to the past, and hence thought the meeting had been re-scheduled for Monday). However, Boroditsky also found that participants who had answered blocks of questions about *spatial* primes also answered the *temporal* target in a schema consistent manner.[13]



The hat-box is in front of the Kleenex. TRUE/FALSE

**Figure 6.11** Object-moving scenario used as a spatial prime in Boroditsky (1998).



The flower is in front of me. TRUE/FALSE

**Figure 6.12** Ego-moving scenario used as a spatial prime in Boroditsky (1998).

These findings were backed up by reaction time studies that showed that participants were slower in answering within domain questions if they had just answered schema-inconsistent questions in that domain (e.g. subjects were slower to answer say an *ego-*

[13] Boroditsky did not find any reverse priming from time $\Rightarrow$ space. The implications of this will be discussed later.

*moving* metaphor phrased question after a *time-moving* metaphor prime than after an *ego-moving* metaphor prime in both space $\Rightarrow$ space and time $\Rightarrow$ time questions) and also about temporal relations if they had just answered schema-inconsistent questions about spatial relations. From these results, it would appear that structured on-line mappings occurred between spatial and temporal information in answering the target questions.

To return to the amended version of the framework described by Murphy (1996), Boroditsky's findings are consistent with both the Strong Mental Representation and the Weak Mental Representation 2 views, and would appear to provide some empirical support for the existence of structural (analogical) priming between the representations of real world concepts associated with real world lexical items (as opposed to the invented dictionary definitions in experiment 4).

### 6.8.4 Experiment 5

I noted earlier two potentially damning objections to the findings of experiment 4:

1. Firstly, the "dictionary entries" in the main task were artificial, indeed little more than definitions, and therefore not necessarily cognitively plausible.

2. Secondly, participants were presented with the tasks on paper, and had unlimited time in which to solve the inferencing problems and reconcile and map any "meanings" of words in the various base and target specifications. Since structure mapping is a computationally expensive process - especially in comparison to mapping surface features - experiment 4 has little relevance to the on-line demand characteristics of analogical processing "in the wild".

I noted that what was needed in order to address these concerns, were two concepts that both supported contrary systems of inferential support from a single lexical term (to mimic the reversal of inference patterns prompted by the materials in experiment 4) *and* supported interpretations of one another analogically (to capture the structural determination of mappings between dictionary items in experiment 4). From the foregoing, it would appear that space and time meet these requirements perfectly. Both support contrary systems of inferential support from a single lexical term - "move forward" can yield contrary interpretations in both domains - and, as Boroditsky (1998) has shown, spatial primes can support temporal interpretations of these terms analogically.

Experiment 5 was designed to use these factors to meet these objections, and show that the results of experiment 4 could be replicated, with participants - using their own internal representations of spatial and temporal concepts - solving "analogical" problems in working memory whilst using parallel on-line structure mappings to map between lexical items in the representation s of the analogies.

#### 6.8.4.1 Participants

The participants were 85 volunteers, a mixture of postgraduate and undergraduate students from the Department of Artificial Intelligence, Centre for Cognitive Science, Department of Psychology and the Faculty of Music at the University of Edinburgh.

#### 6.8.4.2 Materials and Design

The basic analogical problem used is shown in figure 6.13. A base scenario and two targets were constructed. In order to generate an analogical inference for the base, participants would have to map it to one of the targets, and then infer one of two contrasting inferences, each of which received structural support from one or other of the targets. However, in the base, one key structural element was ambiguous (moved forward temporally), and could reasonably be mapped onto structural elements in either of the targets (it could mean something was going to happen earlier in the week or later in the week). Participants could map "moved forward" temporally to the early part of the week, "Monday" or "Tuesday", or the later part of the week, "Thursday" or "Friday" in the scenarios. The former mapping would in turn support the mapping

CAUSE(VISIT_EARLY_IN_WEEK(MacCAWBER, BUSINESS),
CARRY_OUT(MacCAWBER, UNPLEASANT_ACTION),

leading participants to carry the

CARRY_OUT(MacCAWBER, UNPLEASANT_ACTION)

inference to the base, whilst the latter mapping ("Thursday" or "Friday") would support the mapping

CAUSE(VISIT_LATE_IN_WEEK(MacCAWBER, BUSINESS),
CARRY_OUT(MacCAWBER, PLEASANT_ACTION),

leading participants to carry the

CARRY_OUT(MacCAWBER, PLEASANT_ACTION)

inference to the base.

In order to facilitate the parallel analogical mappings, prior to making their inferences participants answered three TRUE/FALSE priming questions about spatial scenarios. The scenarios used either the ego-moving frame of reference (see Figure 6.12), or the object-moving frame of reference (see Figure 6.11). It was predicted that the two frames of reference would map onto (and bias the use of) the ego-moving and time-moving perspectives in time, respectively, influencing the way that "moved forward" was mapped in the larger analogy task.

The task was set up to run on a web browser. The basic analogy stimuli were designed so that the analogous target (Marge) was always consistent with the spatial prime. On the first page, participants were presented with the base and targets and some test questions relating to them. There the followed some "diversion tasks" - a page of unrelated tasks, and then the primes. On a separate page that immediately followed the primes, participants were told: "Angela told Geoff 'next Wednesday's meeting has been moved forward two days.'", and then asked to indicate what the implications of this were from the initial analogy: What did she [Angela] intend Geoff to guess?" All participants also provided a confidence score for their answer to the target question on a scale of 1 to 5 (1=not at all confident, 5=very confident). On the following page, participants were asked to indicate which day the meeting had been rescheduled for, and again provided a confidence score for their answer to the target question on the same scale.

### 6.8.5 Procedure

Participants completed the on line questionnaire individually with no time restrictions. The order of presentation of the targets was randomised. Each participants received either the *ego-moving* prime diversion task or the *object-moving* prime diversion task. One control group of participants responded to the target sentences, however the diversion task they were given did not contain the primes. Since the structural mapping at the analogy level in this task is determined by the interpretation of "moved forwards", and since - unlike in experiment 4 - participants would assign an interpretation to the unprimed term, it was expected that the overall pattern of inferences in this group would conform to that of a further control group that was simply posed the question "next Wednesday's meeting has been moved forward two days. What day is it on now that it has been re-arranged?" (in experiment 4, participants had no prior knowledge of the novel term; here it was clear that "moved forward" would be interpreted one way or another, as described above).

**BASE**

Geoff runs a coffee shop in the business district of Nawatobee, a small city in the American mid-west. The coffee shop is called MacCawber's. It is owned by Mr Fingus MacCawber, a fierce fastidious man who owns six other businesses in the greater Nawatobee area. Mr MacCawber visits each of his businesses for precisely 45 minutes on the first Wednesday of every month, preferring to let his managers manage for the rest of the time.

One day Angela, Mr MacCawber's secretary rang Geoff with what she said was some important news. However, Angela told Geoff that she couldn't tell him the news, as that would be a betrayal of confidence, so she would just have to provide him with some clues, and leave him to guess.

**TARGET 1 (Many surface commonalities with base)**

"Well Geoff," said Angela, "Do you remember Birt, who used to run an ice-cream parlour, also called MacCawber's, that was also in the business district of Nawatobee? Last December, Mr MacCawber visited on a day other than usual, going to the ice-cream parlour on the Tuesday, and he sacked Birt for incompetence. Mr MacCawber was heard to say 'I always like to deal with unpleasant business on Mondays and Tuesdays - it gets me in a mean mood for the rest of the week!'

**TARGET 2 (Few surface commonalities with base)**

"Secondly, you may be familiar with the case of Marge. Marge manages MacCawber's large car-part dealership - called Part-U-Need - in the suburbs. In January, Mr MacCawber altered his visit to her, going to the dealership on the Thursday instead of the Wednesday. Marge had been given a big raise. Mr MacCawber had been heard to remark I always like to deal with pleasant business on a Thursday or a Friday - it sets up my mood for the weekend's golfing!"

**Figure 6.13** The basic scenario: used in experiment 5

### 6.8.6 Results

74.4% of participants responded in a schema consistent fashion, excluding data from 2 participants who did not respond correctly to all of the prime questions, and 3 participants whose analogical inference was inconsistent with their indication of the new day for the rescheduled meeting (there were no differences in error rates across conditions). As predicted by the on-line "semantic" structure mapping hypothesis, when given *ego-moving* primes 72.7% of participants inferred that Geoff was going to get a raise (or some other similar positive experience), and confirmed this by saying that Wednesday's meeting had been moved to Friday (the ego-moving consistent response). E.g. they mapped MOVED_FORWARD to VISIT_LATE_IN_WEEK and thus from

CAUSE(VISIT_LATE_IN_WEEK(MacCAWBER, BUSINESS),
CARRY_OUT(MacCAWBER, PLEASANT_ACTION),

in the target they imported CARRY_OUT(MacCAWBER, PLEASANT_ACTION) to the base. When given *object-moving* primes 76.1% of participants inferred that Geoff was going to get fired (or some other similar negative experience), and confirmed this by saying that Wednesday's meeting had been moved to Monday (the ego-moving consistent response). E.g. they mapped MOVED_FORWARD to VISIT_EARLY_IN_WEEK and thus from

CAUSE(VISIT_LATE_IN_WEEK(MacCAWBER, BUSINESS),
CARRY_OUT(MacCAWBER, UNPLEASANT_ACTION),

in the target they imported CARRY_OUT(MacCAWBER, UNPLEASANT_ACTION) to the base.

A 2x2 chi-square analysis showed the prime consistency bias to be significant ; $\chi$ (1, N=43) = 10.29, p<.0.001. This schema-consistency effect suggests that there was on-line analogical transfer from the spatial primes to the temporal lexical items in the representation of the basic analogy, and that these mappings were used in parallel in computing a global structure mapping to generate an inference in the base.

In the control condition where the basic task was identical, but participants were not given spatial primes, responses were fairly evenly split, with 55% of respondents making the early in the week inference and re-assigning the meeting to Monday, and 45% of respondents making the late in the week inference and re-assigning the meeting to Friday (agreement between meeting re-assignment and subsequent inference was

100%). This was consistent with the proportion of participant responses in the control condition where participants were simply given the straight "Wednesday's meeting has been moved forward" task. Here participants understood the meeting to have moved to either Monday or Friday equally (50%:50%). Furthermore, no bias was shown towards surface features (unlike in experiment 4), with 45% of respondents' inferences resulting from an alignment with the analogous target, and 55% of respondents' inferences resulting from an alignment with the literally similar target.

Though this finding appears at first to be inconsistent with the finding of a similar control in experiment 4, it should be noted that in experiment 4, participants had no prior knowledge of the novel term, and so their interpretation of it could not add structural support to either of the target analogs (thus participants used surface commonalities as their means of determining similarity between the base and a target). Here it was clear that "moved forward" would be interpreted one way or another, and that this interpretation would lend structural support to one inference or another, as these findings suggest: consistent with the overall hypothesis that structure would be the prime determinant of similarity in the task, it appears that the structural support prompted by participants unprimed interpretation of "moved forward" has simply overridden any surface feature biases for one target over another.

Thus this finding is consistent with the results of Experiment 4, and corroborates the hypothesis that people can use structure mapping to map "semantic items" in the representations of analogical problems in parallel to using structure mapping to resolve the analogical problems themselves.

No significant differences were noted in mean confidence ratings for primed inferences or day assignments versus unprimed inferences or day assignments. However, interestingly, t tests showed that in both of the primed conditions, the confidence ratings for all (correct and incorrect) inferences were significantly lower than the mean confidence ratings for the inferences that depended on them (*ego*, $t=2.191$ $p<.0.05$; *object*, $t= 2.371$ $p<0.05$). In the control condition this difference was not significant (although the trend was repeated).

### 6.8.7 Discussion

This experiment further examined the hypothesis put forward in the first part of this chapter, namely that the supervenient nature of analogy and categorisation makes the inclusion of 'semantic links' inappropriate at the level of description generally posited by theories of analogy. The hypothesis tested was that, consistent with the claim that analogical and conceptual mappings cannot be distinguished at a cognitive process level and the results obtained in experiment 4, participants would use the same process that they used to process and make analogical inferences to process and reconcile semantic discrepancies they encountered in the representations of base and target analogs.

These findings add further support to the no-distinction hypothesis. As cast by Holyoak and Thagard, the question of semantics revolves around supplying an account of what happens when two semantically similar terms are encountered during the mapping process. In ordinary usage, the representations of human category information involved in these processes are implicit - people have compiled their knowledge of what EARLIER, LATER and MOVED_FORWARD mean, and it is difficult to get a handle on how they use this knowledge to reconcile semantic differences between the two terms. However, whilst the *exact* nature of this compiled knowledge, and the way it is then represented, are at present inaccessible, we have shown that it is possible to specify and model the underlying process involved in the 'semantic reconciliation' of two lexical items in controlled conditions. Asking participants to make inferences with the aid of two targets, and through controlled variance of information representing the semantics of the term that was crucial to determining the representation of higher order structure in the base and each of the targets, enabled the representations participants used for semantically reconciling particular terms during their analogising to be controlled.

The prediction that participants would use the same mapping process to match semantic items in their representations as they would in ultimately determining their analogies was supported emphatically by the results of experiment 4, which used explicit, artificial stimuli, and by experiment 5, which exploited known structure in participants' own internal representations of terms and their possible interpretations.

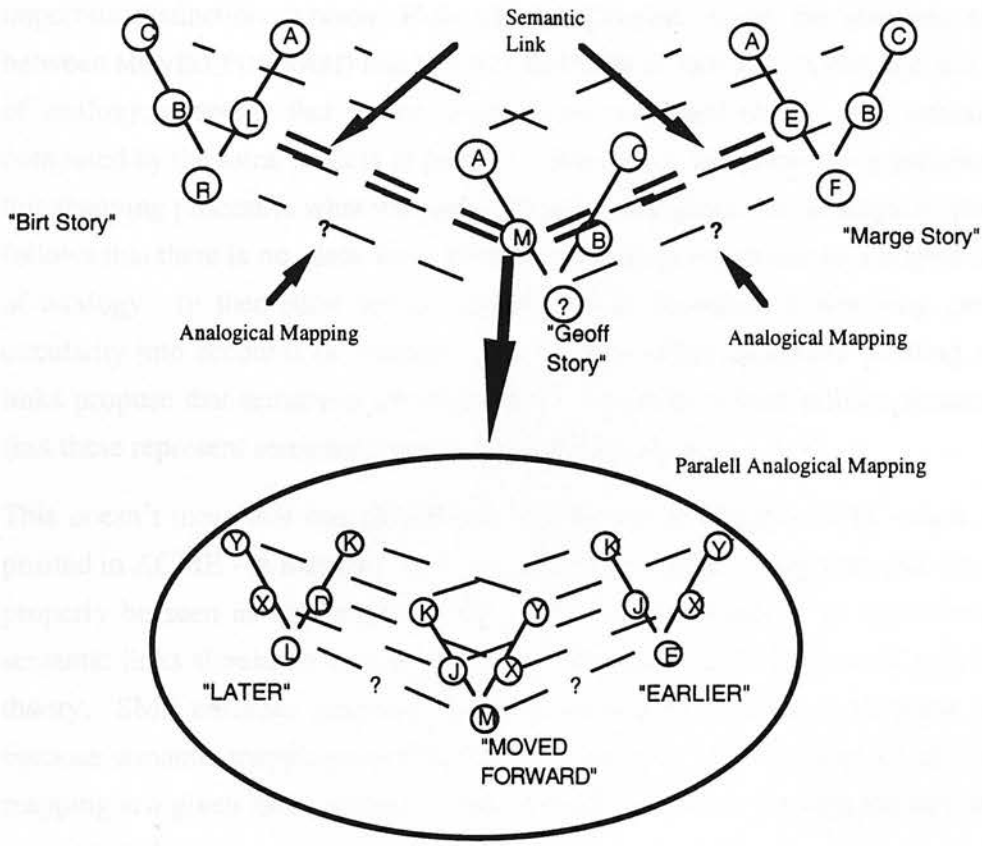



Figure 6.14: **Two views of the analogical mapping process revisited**

## *6.9 Semantic links?*

From the perspective argued for in this thesis, the mapping process used by participants in experiments 4 and 5 is what is meant when one talks about the 'analogical' process. As I argued earlier, it follows from this that I can see no room for a semantic similarity constraint in a cognitive theory of analogy (or at least, it follows that before one can posit such a constraint, one will have to successfully distinguish 'analogy' from 'categorisation').

If we revise the schematic representation of the mapping process presented in figure 6.14 above, in ACME, Holyoak and Thagard model the mapping of E (EARLIER) versus L (LATER) onto M (MOVED_FORWARD) in the making of an analogical mapping between BIRT and either MARGE or GEOFF by means of a semantic link, whereas the evidence presented above points towards this mapping being made by a similar parallel process, operating at another level of conceptual description. This is an important distinction: whereas Holyoak and Thagard model the semantic similarity between MOVED_FORWARD and MOVED_EARLIER or MOVED_LATER as a sub-process of analogy, it seems that in the experiments presented above, this similarity was computed by the same process in parallel. Since from the perspective presented here, this mapping process is what we mean when we talk about the 'analogical' process, it follows that there is no place for a semantic similarity constraint in a cognitive theory of analogy. In theoretical terms, semantic links introduce a worrying element of circularity into accounts of analogy. As I noted earlier, accounts positing semantic links propose that semantics are mapped by a process which utilises semantic links (but these represent semantic mappings in the first place).

This doesn't mean that one should rule out the use of semantic links - such as those posited in ACME - in mapping networks altogether: rather, I suggests that they should properly be seen as implementational details within a model. In modelling terms, semantic links should be seen as notation, rather than embodiments of psychological theory. SME enforces mappings between predicates with the same name precisely because semantic mappings are viewed as external to the workings of an individual mapping at a given level; semantic links should be seen as a notational device for the same procedure.

## 6.10 Conceptual Metaphors?

One important possible objection to the account of analogy, categorisation and semantics presented in this thesis should be considered. I noted earlier that Boroditsky's (1998) study showed that although spatial primes influenced subsequent temporal reasoning (effects replicated in experiment 5), there was no reverse priming effect from time to space. Lakoff and Johnson (1999) interpret these results as lending support for a strong metaphorical theory of representation such as that proposed by Lakoff and colleagues (Lakoff and Johnson, 1980; Lakoff, 1987a; Lakoff and Johnson, 1999).

According to Lakoff and Johnson's Conceptual Metaphor theory the human conceptual system is structured around a small set of experiential concepts. This small set of concepts emerges directly out of experience and its members are defined "in their own terms." Such fundamental concepts include basic spatial relations (up/down, front/back), a set of "physical ontological" concepts (entity, container), and a set of basic experiences or actions (eating, moving). The Conceptual Metaphor theory grounds all other concepts in these basic experiential concepts, proposing that concepts which do not emerge directly out of physical experience must be metaphoric in nature, and that these "abstract" concepts are represented via on-line mappings to the fundamental experiential concepts (Lakoff and Johnson, 1980; Lakoff, 1987a; Lakoff and Johnson, 1999). Thus, to return to Murphy's (1996) classification, Conceptual Metaphor theory equates to Strong Metaphoric Representation, in that abstract concepts are not understood via their own representations, but instead by metaphoric reference to another domain; abstract concepts are represented only by links to concepts that *are* actually represented as entities. This view would propose that *space* concepts are represented independently, and *time* concepts are represented by links to *space* concepts.

Clearly the space to time mapping found by Boroditsky (1998) and the results of experiment 5 provide support for some kind of on-line metaphoric mapping. This evidence is therefore at odds with Murphy's (1996) Structural Similarity and Weak Metaphoric Representation (WMR1) hypotheses, described in section 6.8.2, which rule out the possibility of on-line mappings. The question I wish to consider here is whether this evidence best supports, and is best explained by either the Strong Metaphoric Representation (or Conceptual Metaphor) hypothesis, or the second Weak Metaphoric Representation (WMR2) hypothesis I described in chapter 6.

Evidence for Strong Metaphoric Representation might appear to point the way to a simple, ready made answer to the problem of characterising cognitive conceptual skills that this thesis has grappled with: Conceptual Metaphor theory. Below, consistent with the main thrust of argument thus far in this thesis, I shall argue that such simple appearances can be deceptive.

### 6.10.1 Two Types Of Metaphoric Representation

To briefly summarise the two positions from section 6.8.2:

According to strong MR, concepts are not understood via their own representations, but instead by metaphoric reference to another domain (i.e. represented only by links to concepts that *are* actually represented as entities). According to this view, *space* concepts are represented independently, and *time* concepts are represented by links to *space* concepts.

The Weak MR2 hypothesis posits that that some parts of conceptual metaphors are metaphoric in the manner of Strong MR, whilst other parts may be reflective of prior structural mappings that have become conventionalised, or aspects of the actual concept itself. According to this view, *space* and *time* concepts are represented independently, but there may be metaphoric links between them in either direction.

### 6.10.2 The Conceptual Metaphor Theory Account Of Time Concepts

According to Conceptual Metaphor theory (Lakoff and Johnson, 1999), it would appear *all* temporal reasoning would *necessarily* involve some kind of on-line alignment - i.e. structure map - between a representation of a temporal construct (say in the interpretation of a temporal phrase) and an appropriate spatial concept. Moreover, Lakoff and Johnson's broader statement of Conceptual Metaphor theory implies that this mapping is necessarily one way: abstract concepts are *not* understood via their own representations, but instead by metaphoric reference to another domain (Lakoff and Johnson, 1980; 1999).

With regards to time, Lakoff and Johnson (1999, p139) state:

> "it is virtually impossible for us to conceptualize time without metaphor... Time is as basic a concept as we have. Yet time in English, and in other languages is, for the most part, not conceptualised and talked about on its own terms. Very little of our understanding of time is purely temporal. Most of our understanding is a metaphorical version of our understanding of motion in space."
>
> Lakoff and Johnson (1999, p139).

> "we cannot measure time-in-itself, whatever that could mean. We can only define *time* to be that which is measured by regular iterated events. Therefore we cannot take the common-sense understanding of time at face value from a cognitive perspective. However... if we start from the view that time is conceptualized through the comparison of events, we can arrive at an adequate analysis of the common-sense understanding of time itself."
> Lakoff and Johnson (1999, p154).

> "Can we avoid... metaphors, and think and talk about time only literally? No. Our conceptual and linguistic systems do not allow it."
> Lakoff and Johnson (1999, p168).

Lakoff and Johnson argue that the spatial metaphor systems used to conceptualise time are the *ego-moving* metaphor (*ego-moving* system) and *object-moving* metaphor (*time-moving* system) described in section 6.8.1. I do not wish to dwell on Lakoff and Johnson's philosophising on time here. The main point is that Lakoff and Johnson make strong claims about the way time is conceptualised, and empirical evidence for the kind of the on-line mapping claims they make is provided by Boroditsky's (1998) asymmetry of priming results.

Apart from Boroditsky's results, Lakoff and Johnson offer a great deal of linguistic evidence, e.g.

> The precious seconds *oozed* through my fingers. The deadline *sneaked* by me. The deadline was *marching* towards me like a brass band. The days *cascaded* by.
> Lakoff and Johnson (1999, p149).

Lakoff and Johnson point out that all these words indicate motion, and all express the idea that time is passing by or towards an observer. However, as Murphy (1996) points out, one should be wary of linguistic evidence when evaluating claims of this kind, as the words used in conceptual metaphors are both the predictor (the linguistic difference) and the data (the cognitive difference). Murphy offers the following parody of the Whorfian position in illustration of this point:

> "Whorfian: Eskimos are greatly influenced by their language in their perception of snow. For example, they have *N* words for snow... whereas English has only one, *snow*. Having all these words makes them think of snow very differently than, say, Americans do'

> "Skeptic: How do you know they think of snow so differently?
>
> "Whorfian: Look at all the words they have for it! *N* of them! They must make a lot of distinctions between kinds of snow that we don't, since we just call it *snow*."
>
> Murphy (1996, p183).

As Murphy notes, in arguments like this, the claim is that language - or linguistic constructs - influence thought, but the evidence given for this is the language or linguistic constructs themselves. Thus Boroditsky's (1998) priming asymmetry - which appears to involve more than purely linguistic evidence, could be an important piece of evidence for Conceptual Metaphor theory.

### 6.10.3 Is Time Dependent Upon Space (Or Other Such Things)?

It is not clear, however, either that Boroditsky's evidence provides support for Strong Metaphoric Representation, or that it is best explained by such an account. There are - at least - two main objections to the idea that Boroditsky's priming results support Strong Metaphoric Representation. Firstly, it isn't clear that the Strong Metaphoric Representation hypothesis - insofar as it can be understandably be interpreted - can sensibly specify a concept of time. And, secondly, following from this, differences in the stimuli Boroditsky used in priming space ⇒ time and time ⇒ space make any direct comparison of results in the two conditions impossible. The asymmetries demonstrated by Boroditsky (1998) could be indicative of this latter fact, and not the influence of Strong Metaphoric Representation as claimed.

To take the matter of the making sense of the Strong Metaphoric Representation of time. As Fillmore (1971) crucially acknowledges, in order to understand the *ego-moving* and *object-moving* metaphors, one has first to have an understanding of time independent of those metaphors:

> "The words "earlier" and "later"... are basic temporal notions, not based on a movement metaphor. In fact, an understanding of the setting of the front/back axis for an object in motion presupposed an understanding of unidirectional time, since "front" was defined in that case in terms of a part of something "arriving earlier" than the rest of it"
>
> Fillmore (1971, p46).

Lakoff and Johnson (1999) seem - to an extent - to be aware of this problem. They ultimately define time - metaphorically - in relation to an idiosyncratic notion of events, and acknowledge that time seems to be essential to understanding these. But Lakoff and Johnson argue that our supposition that time is more basic than events is in fact evidence of the Conceptual Metaphor system. This is because, they say, we see times as locations or containers for events because we are metaphorically conceiving time in terms of spatial locations or containers.

However, this doesn't address the point raised by Fillmore. Understanding "earlier" doesn't appear to depend upon a metaphor (container, spatial or otherwise): it depends upon having grasped a basic temporal concept. If we return to McTaggart's (1908) observation, "If *M* is ever earlier than *N*, it is always earlier. But an event, which is now present, was future, and will be past." McTaggart, (1908 , p 24). This seems as good a statement of a basic temporal concept as any, and it doesn't rely on time "containing" events, or "locating them in space" - it simply points to basic *temporal* relations that hold between them. Indeed, such relationships can hold within an event: an earlier part of an event is earlier than a later part of an event, and that part which is now present, was future, and will be past. Lakoff and Johnson (1999, p138) themselves admit "events have beginnings and ends"; what they don't explain is the point Fillmore raises: how exactly is one is supposed to conceive of these beginnings and ends without a basic notion of time? Lakoff and Johnson (1999, p137) claim that "All of our understandings of time are relative to other concepts such as motion, space and events."[14] Clearly time could not be understood purely in terms of static space. But motion and events are spatio-temporal at the very least (and the spatial point is arguable in the case of mental events). What Lakoff and Johnson singularly fail to specify is how these are supposed to conceptualised independently of a concept of time.[15] (See McTaggart, 1908, for a lucid description of some of the many complexities that the *primâ facie* concept of time brings with it.)

---

[14] Lakoff and Johnson also claim that "motion appears to be primary and time is metaphorically conceptualised in terms of motion. There is an area in the visual systems of our brains dedicated to the detection of motion. There is no such area for the detection of global time." (Lakoff and Johnson, 1999, p140). But, as Fillmore's point is intended to show, motion is an inherently *spatio-temporal* concept.

[15] Lakoff and Johnson (1999, p 176) describe the following event structure: Events have an Initial State, a Start, an End of Start, a Main Process, Possible Interruptions, Possible Continuation or Iteration and Resultant State. This structure is learned metaphorically from "specialised notions" (e.g. self-propelled motion and force). Again, what they fail to describe is how these specialised notions or

### 6.10.4 Priming From "Space To Time"

| | |
|---|---|
| Saturday comes before Friday | **TRUE** or **FALSE** |
| Wednesday comes after Tuesday | **TRUE** or **FALSE** |
| Thursday comes before Wednesday | **TRUE** or **FALSE** |
| Friday comes before Saturday | **TRUE** or **FALSE** |
| Wednesday comes before Thursday | **TRUE** or **FALSE** |

**Figure 6.15** "Time-moving" scenario used as a temporal prime in Boroditsky (1998).

| | |
|---|---|
| On Saturday, Friday is before us | **TRUE** or **FALSE** |
| On Wednesday, Tuesday is behind us | **TRUE** or **FALSE** |
| On Thursday, Wednesday is before us | **TRUE** or **FALSE** |
| On Friday, Saturday is before us | **TRUE** or **FALSE** |
| On Wednesday, Thursday is before us | **TRUE** or **FALSE** |

**Figure 6.16** "Ego-moving" scenario used as a temporal prime in Boroditsky (1998).

The underspecification of the totally metaphoric concept of time is, then, one strand of my argument against interpreting Boroditsky's priming asymmetry findings as evidence for Strong Metaphoric Representation. The other strand of this argument relates to the nature of the tasks participants performed in Boroditsky's experiment.

The *ego-moving* and the *object-moving* metaphoric systems are movement schemas. That is, they are structured according to concepts of objects moving in space (see section 6.8.1). In the space ⇒ time experiments in Boroditsky (1998), this motion

---

concepts are understood *without* a concept of time, which, since the event structure they describe is temporal in the extreme, is a serious lacuna.

was explicit in the primes given to subjects (figures 6.11 and 6.12) in the shape of an arrow indicating movement on a visual prime.

However, in the time ⇒ space experiments, subjects were simply presented with blocks of text questions (figures 6.16 and 6.15).

What is not clear at all, however, is whether these time ⇒ space primes actually involve *any* motion. Clearly, no explicit motion is present in the stimuli. But whether implicit motion is present is a moot point too (see McTaggart, 1908). Consider: both time and space are usually measured numerically[16] (in time or distance units). In counting to ten, it is possible to adopt both the *ego-moving* system and *object(number?)-moving* system. In the former, when one gets to five, six is still ahead, whilst in the latter, one comes before two. Yet it isn't entirely clear that counting in one's head carried any implicit concept of *motion*. And, one might argue, as for *numeric* scales, so for *calendric* ones.

Two points are worth making in respect of this. Firstly, no *evidence* is presented to support the idea that the primes imply motion; and secondly, any positive claim for "motion" in these stimuli seems necessarily to incur the kind of Whorfian circularity described by Murphy (1996, discussed above). The claim that the time stimuli prime motion would presumably appeal to the idea that they use the same *words* that are used to describe spatial movement. I detailed above Murphy's description of the way arguments like this claim that language - or linguistic constructs - influence thought, but then only provide evidence in terms of the language or linguistic constructs themselves. It would appear that the asymmetry shown in Boroditsky's priming experiments relies on just this kind of claim. It is assumed that words like "ahead" and "behind" imply temporal motion because they imply spatial motion, and then the failure of *temporal* "ahead" and "behind" to prime an appropriate *spatial* motion schema is used as evidence for Strong Metaphoric Representation.

Like the Whorfian analysis above, the initial assumption about linguistic stimuli is essential to the interpretation of the linguistic data which provides evidence that the temporal terms are understood in relation to the spatial schema: the initial assumption about the linguistic data.

[16] I'm grateful to Eshaa AlKhalifa for this insight, and for many interesting discussions about time.

McTaggart (1908) presents some good reasons to doubt these assumptions, suggesting that these terms do not necessarily imply motion in this case. McTaggart notes that

> "in order to get change [in a spatial system], and change in a given direction... [one position must]... be in the Present to the exclusion of others, and this characteristic of presentness should pass along the series in such a way that all positions on one side of the Present have been presents, and all positions on the other side of it will be present."
> McTaggart (1908, p 463).

Whilst the graphical representations of the spatial primes clearly establish present, and explicitly define motion, this is not the case with the spatial primes. According to McTaggart's account, no change, and therefore no motion is present in the spatial primes.

### *6.10.5 Space, Time And Weak MR2*

In conclusion then, the Strong Metaphoric Representation hypothesis seems unable to explain coherently why temporal concepts can only be understood metaphorically from spatial ones. Nor does the one piece of empirical support claimed for it (Boroditsky, 1998; see Lakoff and Johnson, 1999) provide any compelling reason to overlook this explanatory lacuna. Other evidence too, supports a skeptical view of Conceptual Metaphor theory. For instance McGlone (1996) in a series of experiments examining metaphor comprehension found little evidence that underlying Conceptual Metaphors had any significant role to play, when it came to paraphrasing metaphors, recalling them, or finding similarities between metaphor meanings. [17]

Thus Conceptual Metaphor theory does not seem to offer any panacea for the problematic view of concepts put forth in this thesis. A dispassionate look at the relationship between space and time seems to tally with the view I have termed WMR2. There is evidence that concepts prime metaphorically (Boroditsky, 1998), but it doesn't seem that this evidence is amenable to explanation in terms of neat, unidirectional "master / slave" metaphoric mappings. There is no evidence that time is

---

[17] McGlone favours the 'attributive categorisation' view of metaphor, whereby metaphors are interpreted as 'class inclusion statements' c.f. Glucksberg and Keysar (1990). Given the problems noted throughout this thesis of saying just what 'class inclusion' amounts to, this seems to some extent to be simply a matter of switching the problem.

understood in purely metaphoric terms, and there are good reasons for believing the contrary.

The evidence indicates some aspects of the *ego-moving* and *time-moving* systems used for sequencing events may have originally been mapped metaphorically from space but they have now become conventionalised. For instance, it is not clear whether *before* or *after* are "spatial" concepts at root anymore, whatever their etymological ancestry. Instead, they seem best understood as metaphors for earlier and later (contra Conceptual Metaphor theory, this might also indicate that metaphoric mappings can run from time $\Rightarrow$ space as well as space $\Rightarrow$ time). Some *time* concepts are - perhaps - no longer connected to *space* concepts via an on-line mapping, even though they may have been shaped by spatial metaphors. On the other hand, other aspects of the *ego-moving* and *time-moving* systems used for sequencing events may still be mapped metaphorically from space on-line, as Boroditsky's results indicate. Once more, the messy view of conceptualisation seems to provide the best fit with the available evidence.

## 6.11 Summary

The preceding chapters presented the case - and some evidence for - the 'no distinction' hypothesis, which denies any simple distinction between analogy and categorisation at the cognitive process level. In this chapter, I have explored some of the theoretical implications the 'no distinction' hypothesis has for existing theoretical approaches to analogy. The two experiments presented have shown how the removal of any neatly bounded 'concept domains' with which various aspects of the analogical process can interact has major implications for theories based on just such assumptions, and in particular for the semantic similarity constraints which are variously included in, or excluded from, theories and models of the analogical process.

In chapter 4, I noted that structure may have a large part to play in the determining of conceptual mappings. In these experiments, it was the mapping of structures - the process that from a cognitive science point of view is ordinarily seen as definitive of analogy - that determined the way in which semantics were reconciled. I have argued that theoretically these results seem to indicate that including semantic links into an account of the process that ultimately determined semantic linkage introduce a worrying element of circularity into accounts of analogy, and that as a result, the semantic links proposed within models like ACME should not be seen as embodiments of psychological theory.

All theories of analogy make some reference to conceptual semantics, either internally, by use of some kind of semantic links - as described above - in order to reconcile semantically similar terms (Holyoak and Thagard, 1995; Keane, Ledgeway and Duff, 1994; Hofstadter, 1995), or else externally, by reference to some externally defined canonical conceptual representation (Forbus, Gentner, Markman and Ferguson, 1998). From a clarity point of view, it might be argued the latter approach is to be preferred, since it explicitly leaves the problem outside of the scope of the model (although, as I noted earlier, this rules out the possibility of simply importing structure mapping theory to explain similarity *within* categorisation judgements). The fact that SME requires that predicates which are to be mapped be lexically identical does at least highlight the fact that analogical models are underspecified in this area, whereas the way that models like ACME *apparently* resolve semantic ambiguities (although these are in fact solved by the programmer setting the network) might be seen as obscuring this important issue. In the end, though, these are mainly matters of stylistic preference. A more interesting question to consider is whether the results and analysis presented so far indicate that there may be a limit to the extent to which all mapping models that make these external appeals can ultimately cope with the complex, parallel nature of human analogy making.

# Chapter 7

# Discussion And Conclusions

## 7.1 Introduction

This chapter discusses the analysis and results described in this thesis. The key findings of the theoretical survey presented in Chapters 2, 3 and 4 are summarised, and then related to the findings of the empirical studies in Chapters 5 and 6. Some of the limitations of these studies, and the general approach adopted, are identified.

In the light of the 'no distinction' hypothesis presented, a speculative framework for discussing modes of conceptualisation is outlined, and some potentially fruitful future research directions are described.

## 7.2 Summary Of Findings

### 7.2.1 Analogy And Categorisation

The exploration of the links between analogy and categorisation in this thesis has involved a re-appraisal of the basic assumptions made about concepts in psychology, and a systematic evaluation of the theoretical underpinnings of three decades of categorisation research.

At the head of this work was the rigorous re-analysis of the theory of concepts attributed to Wittgenstein (1953). This demonstrated that the characterisation of Wittgenstein's views that has held sway in the cognitive psychology literature (c.f. Rosch, 1978; Johnson-Laird, 1983; Medin and Ortony, 1989; Lakoff, 1987) is at considerable variance to a close - but straightforward - reading of all of the relevant sections of Wittgenstein (1953).

The empirical review that followed this demonstrated that:

- little of the research that claims its inspiration from Wittgenstein bears much relation to his views [1]
- the picture that emerges from the interpretation of Wittgenstein presented is consistent with many of the findings of psychological studies of concepts; moreover, this picture can provide the conceptual glue required to make sense of these findings.

To briefly summarise this position and its relation to the evidence, Wittgenstein argues:

- that categories have no necessary or sufficient defining characteristics: rather that "family resemblances" can be traced across categories (supporting evidence is summarised in Smith and Medin, 1981)
- that these category spaces are unbounded - i.e. there are no boundaries to the space across which "family resemblances" can be traced (Labov, 1973; McClosky and Glucksberg, 1978; Sloman, Malt, Shi, Gennari and Wang 1997 a, b)
- that learning a category such as game does not involve extracting an essence or schema from instances. (Malt, 1994; see also experiment 2)
- that learning a category or concept involves learning examples (instances) of the usage of that concept and appropriate ways of using these examples (Nosofsky 1986, 1991; Lamberts, 1996).

This theoretical and empirical evidence indicates that *conceptual storage* is best viewed not as a question of instances versus generalisations, but rather one of unitary versus multiple representation accounts of concepts. Unitary accounts of categorisation posit a single stored representation - schema, prototype or, perhaps, theory - which is used to determine category membership. Multiple representation accounts posit the storage of a number of representations which may jointly or individually result in some object being categorised as the outcome of some process. These different models of conceptual representation have different implications for theories of categorisation. If a unitary representation model is correct, then one would expect that provided one could specify the stored representation and the process by which objects were related to it, one should in principle be able to give a definitive account of, say, why it is some

[1] Put simply, Wittgenstein did not come to praise concepts, he came to bury them.

things are X's. On the other hand, a multi-representational account would not admit any definitive account of X's at all, since the multiplicity of relations between the differing stored representations that could influence an object's X-ness preclude any kind of general account covering all instances of that concept. A specific instance's X-ness would be dependent upon that - and only that - instance's interaction with some subset of the stored elements relating to the representation of X-ness, and the concomitant process by which X-ness is adjudged.

I argue that this latter view, which is strongly supported by the evidence, is incompatible with the reified view of concepts which dominates current research in cognitive psychology, and is used to justify a two-process account of literal versus non-literal conceptual processing.

### *7.2.2 Analogy And Non-Literal Concepts*

The second strand of work in this thesis looked at analogy and metaphor, considered as theoretical constructs, in relation to the results presented above. It is becoming more widely accepted that relational structure plays an important role in conceptual understanding (Keil, 1989; Medin and Ortony, 1989 Goldstone, 1994): analogical reasoning research has directly addressed and modelled a process which reasons amongst structural networks (Forbus, Gentner and Law, 1995; Holyoak and Thagard, 1995). Consistent with the approach described above, the empirical and theoretical work in the latter part of the thesis attempted to demonstrate that the distinctions used to separate analogy and categorisation research - which characterise analogies as "connections that go beyond ordinary conceptual structures" - are not only unsupported by current evidence, but that they cannot withstand empirical scrutiny.

In the first of these experiments, subjects were given Gentner's classic "Karla the Hawk" stimuli and asked to categorise them. Results showed that participants analogical and categorisation responses to these materials could be explained and predicted by the same process model. A second experiment showed that although subjects had stored the categories they formed, it appeared that they had not abstracted a general representation for them. The third experiment re-ran the first experiment, replacing the stories with descriptions of novel objects, and duplicated the earlier result. Thus in both of the classification studies subjects' performance in categorisation tasks was consistent with their performance in similar analogy tasks, and could be predicted and described by analogical process models.

The claim made in this work is not simply that "analogy and categorisation are identical." In chapter 6 I have tried to show that the relationship between the two is complex, and that analogy theories (Holyoak and Thagard, 1995; Forbus, Gentner and Law, 1995; Bowdle and Gentner, 1998) incorporate assumptions about concepts that cannot be sustained outside of the laboratory. Experiment 4 explored the role of semantics in analogy, in order to expose circularities in current definitions of analogical theories. By presenting subjects with analogy materials containing novel terms, and then presenting characterisations of the use of these terms in various forms, I have shown that semantic matching in analogy can involve exactly the same process as a global computation of an analogy itself, rather than separate processes, as current analogy models assume. In my studies, subjects used "analogy" to map the novel term characterisations and disambiguate semantics in the representations of base and target analogies, before using these mappings to generate further mappings in order to make an analogical inference.

Experiment 5 then used materials developed by Boroditsky (1998) to replicate these effects without recourse to artificial terms and definitions. Boroditsky showed how priming in the spatial domain could influence reasoning about temporal questions. Experiment 5 examined the effect such spatial priming has on analogies involving time, in order to show that in such cases analogical influences can be determined by such priming (which in analogy theory terms would be classified as semantic) rather than the structural means many analogy theories predict (analogy theories seem to explicitly rule out the kind of purely "semantic" analogising carried out by subjects in these studies).

This work has had two motivating concerns. The first is to attempt to strip away some fairly ingrained assumptions and clarify the nature of concepts from a scientific perspective. The two-process account is largely made up of implicit assumptions. It is not a scientific hypothesis. In arguing that these assumptions are unwarranted, and presenting the 'no-distinction' hypothesis, my aim has been to propose an account of conceptualisation that fits the available evidence, and put forward this account in terms of a hypothesis that is empirically testable. The statement that there is no clear, principled distinction to be made between analogical and 'straight' categorical processing is clearly falsifiable. And falsification of the 'no-distinction' hypothesis might pave the way to a clear, principled statement of a two-process account, if such an account is justified.

The second motivation was more pragmatic. This work ultimately seeks to show not only how a reified view of concepts makes analogy models brittle once one pursues conceptual representation beyond a fixed level of description, but also how a richer understanding of analogy can illuminate our understanding of concepts, In the following sections, I shall consider some objections to the first of these factors, and attempt to show how the second goal might be realised in the future.

## 7.3 The What Of Analogy And Categorisation...

The discussion of Conceptual Metaphor theory at the end of Chapter 6 serves to further illustrate the entangled view of literal and non-literal conceptual processing put forward in this thesis. The view presented here does, however, sit ill with our intuitions about the nature of these processes, with current orthodoxy in analogy, metaphor and categorisation research, and, it would appear, with any idea of formulating a tractable account of conceptualisation. The second of these points has been dealt with at length in this thesis. Here, I wish to consider the former point, and in the section following, I shall discuss the latter.

To deal with the "what" of analogy and categorisation first. The outcome of the investigation presented here, of the interpretation of Wittgenstein's position and the supporting evidence from categorisation research, as well as the empirical studies, is a strong statement of the 'no-distinction' hypothesis. And whilst the 'no-distinction' hypothesis might appear counterintuitive, it need not necessarily be so.

The 'no-distinction' hypothesis amounts to the claim that there is no principled distinction to be made between literal and metaphoric meaning. This appears to leave an important question unanswered: how can we explain why people can often judge a meaning to be literal or metaphoric? To return to the example from French (1996), discussed in chapter 5: what lies behind the intuition that "an orange crate is an orange crate is an orange crate"?

Certainly the reasons behind these intuitions need exploring, however, the ability of listeners to adjudge meanings to be literal or metaphoric need not indicate that listeners *process* so called literal and metaphoric utterances differently. As Rumelhart (1979) has argued, "the classification of an utterance as to whether it involves literal or metaphoric meanings is analogous to our judgement as to whether a bit of language is formal or informal. It is a judgement that can be reliably made, but not one which signals fundamentally different comprehension processes" (p. 79).

Gibbs (1984) argues that one reason why some sentences seem so literal is that listeners are influenced by the interpretative context in which such judgements are made: people judge a sentence as having literal meaning because it is isomorphic with the situation in which the sentence is interpreted (Fish, 1980). However, it doesn't follow from this that the literal meanings of sentences can be uniquely determined, as our understandings of situations *always* influence our understandings of sentences. Says Gibbs, "To speak of a sentence's literal meaning is to already have read it in the light of some purpose, to have engaged in an interpretation. What often appears to be the literal meaning of a sentence is just an occasion-specific meaning where context is so widely shared that there doesn't seem to be a context at all." Gibbs, 1984, p. 296; As for judging whether sentences are literal, I claim, so for judging whether whales are mammals or fish; or, for that matter, whether our picnic is 'on the orange-crate' or 'on the table'.

However, I should note that these arguments try to explain away the intuition problem, rather than explain it. Whilst they offer some explanation of how reliable literal / non-literal judgements could be *reconciled* with the 'no distinction' hypothesis, this account is by no means conclusive.

### 7.3.1 The What Of Analogy

The 'no distinction' hypothesis has clear implications for theories of analogy. As I noted at the end of Chapter 4, all theories of analogy make some reference to conceptual semantics, either internally, by use of some kind of semantic links - as described above - in order to reconcile semantically similar terms (Holyoak and Thagard, 1995; Keane, Ledgeway and Duff, 1994; Hofstadter, 1995), or else externally, by reference to some externally defined canonical conceptual representation (Forbus, Gentner, Markman and Ferguson, 1998).[2] It could appear that in questioning the status of these, I am making little more than an arcane point about which aspect of a given analogical theory is to be preferred. This is not so. The implications of this

[2] One interesting recent model of analogy to note in this respect is LISA (Hummel and Holyoak, 1997). At present, LISA implements the multi-constraint theory of analogy, making reference to explicit semanitc links based upon a taxonomic hierarchy in a similar way to ACME. However, the underlying architecture of LISA appears to offer the potential to implement a more flexible model of conceptual sematics than LISA currently embodies.

analysis for all theories of analogy are far-reaching. The problem of semantics in analogical mapping is indicative of a deeper question. Do the results and analysis presented in this thesis indicate that there may be a limit to the extent to which all mapping models that make these external appeals can ultimately cope with the complex, parallel nature of human analogy making?

To take the matter of the nature of semantics first. Forbus *et al* (1998) attempt to use pre-defined canonical conceptual representations to explain how structure mapping can recursively map non-identical predicates by re-representing any non-matching predicates and then permitting a partial match to be made. Consistent with the analysis of a recursive semantic reconciliation process presented in chapter 6, Forbus et al argue that semantic terms are decomposed into sub-predicate re-representations, and then mappings between these are determined using the same process as the original analogical mapping:

> "re-representation allows relational identicality to arise out of the analogical alignment, rather than acting as a strict constraint on the input descriptions" (Forbus, Gentner, Markman and Ferguson, 1998, p. 246).

So far, these proposals are little more than speculative, and thus, it seems rather unfair to subject them to any rigorous critiquing. However, even in their speculative form, there are important questions that Forbus *et al*'s speculations fail to address, most obviously and most importantly, the question of: where does it all end?

In experiments 4 and 5, I concentrated on the semantic reconciliation of *one* set of similar-but-not-identical terms, and followed this reconciliation process down through *one* level of recursion, where the results indicated - consistent with the 'no distinction' hypothesis and Forbus, Gentner, Markman and Ferguson's (1998) speculations - that the same process was used to resolve semantic ambiguities as was used to determine analogical mappings.

However, it seems unlikely that realistic representations of real-world analogies will contain such a small number of similar yet non-identical predicate matches to reconcile. Real-world representations could, and probably will, contain many more such predicates, and the re-representations of these predicates - whose predicates will also need to be matched during the parallel semantic reconciliation of the original predicates - may contain more non-identical but semantically similar predicates, potentially as a factorial of the original number of semantically similar predicates selected for reconciliation.

Logically, at least, this seems to point to both a combinatorial explosion - in terms of the number of predicates to be reconciled, and hence individual semantic reconciliation sub-processes to be run - and a potential infinite regress: if an *identical* mapping process is to be run recursively, and if re-representation doesn't ultimately uncover *identical* predicate-decomposition representations at some level, then mapping may not terminate. (It might appear that the 'no distinction' hypothesis is a theoretical recipe for disaster.)

However, whilst logic may suggest this, practice suggests that the truth may lie elsewhere. To paraphrase two of Wittgenstein's (1953) more famous ideas, 'words can only decompose into other words' and 'explanations must end somewhere'. The very fact that people can, cognitively, reconcile non-identical but semantically similar terms in real-world situations suggests that this logical conclusion is wrong; something must be wrong with the analysis and characterisation of this process as it stands.

The question this poses is, then: in what way is this characterisation - which seems to be consistent with the empirical data presented above - wrong? How is it that humans can, cognitively, deal with this potential combinatorial explosion, and terminate the recursive search for a mapping?

## 7.4 The Where For Analogy And Categorisation...

As I suggest above, the few existing proposals for describing how this is achieved are largely speculative and incomplete (this stems at least in part from the fact that hitherto there has been a severe lack of empirical evidence that relates directly to this question). Forbus, Gentner, Markman and Ferguson (1998) describe how acculturation and language learning could possibly equip members of a particular linguistic community with representations of domains in terms of canonical sets of dimensions, though what these domains are, and how these dimensions operate in terms of any worked out process of semantic reconciliation is left open: there is nothing in Forbus *et al*'s account that says how, when or why a particular semantic reconciliation sub-process might be invoked, nor ultimately, how recursive mappings in pursuit of semantic reconciliation would terminate.

### *7.4.1 Future Work - A Where For Conceptualisation?*

A possible answer to these questions that could be pursued in the future is the notion of 'semantic focus': the idea that the subprocesses that determine 'semantic similarity' - in the manner discussed so far - will terminate with more or less perfect matches depending on the importance of the support they lend to the focus of a given comparison, i.e. their distance in comparison space from the representations that are the actual point of a comparison. The basic idea at the heart of semantic focus is that semantic mappings will get more diffuse - conceptually weaker, or less rich - the further one moves from the focus of a given comparison. In this section, I will try develop this idea, and sketch a framework in which a two-stage model of analogy and categorisation might be developed.

Whilst I have argued strongly against the idea of canonical conceptual representations *per se* throughout this thesis, I have also noted the evidence from such things as typicality studies that humans do store and use some form of stereotypical, generalised (i.e. generalisations from examples) category information, even though this information is generally insufficient for the determination of categorical similarities and judgements (Rosch, 1978; Komatsu, 1992).

One hypothesis that one might extract from this is that whilst such knowledge may be insufficient when it comes to determining categorisation judgements - i.e. generalisations cannot support mappings that are the focus of comparisons - it may be sufficient to support lower-level matches that act in support of such judgements. To give an example, the claim would be that SMALLER_THAN and LESS_MASS_THAN may be seen as different - even importantly dissimilar - if they are the focus of a comparison, but sufficiently similar if they are merely supporting a particular semantic reconciliation with a larger global comparison of something else, and that similarity at this level might be determined in relation to some pre-computed-and-stored generalisation.

In categorisation terms, a semantic focus hypothesis might appear to be an attempt to "have my cake and eat it." However, I shall argue that this idea can be consistent with the 'anti-schema' position previously argued for in this thesis, in that there is a difference between positing some kind of generalised information, and claiming that this information comes in the form of a schema that is in any way complete or definitive. Rather the claim is that such information represents at best a partial - though pragmatically useful - characterisation of a given concept.

### 7.4.2 A (Very) Speculative Two-Stage Model Of Semantics

In the light of the evidence presented in this thesis, this begs two questions: what kind of representations are these partial representations supposed to be, and how do they interact with the process of determining structural mappings?

To take the former question first: a number of recent approaches to meaning in psycholinguistics have adopted an approach that has some proximity to the exemplar models discussed in chapter 4. In these approaches, vectored similarity matrices have been used to perform similarity analyses amongst the meanings of words in large corpora. I am not going to argue that these models can provide an answer to the conceptualisation problems described in this thesis (and if I believed they did, this would have been a very different dissertation). In fact, I shall argue that they cannot provide such an answer. However, what I shall suggest below is they might suggest an answer to the nature of generalised conceptual similarities, and may provide part of a two-stage model of focused - or diffused - semantics.

I shall begin by briefly describing two leading models in this area: the Hyperspace Approach to Language model (HAL), and the Latent Semantic Analysis Model (LSA).

#### 7.4.2.1 HAL

In HAL theory (Burgess and Lund, 1997), word meanings (or rather similarities between word meanings) are derived from a dimensional analysis of words in context. A corpus of some 300 million words has been analysed by use of a moving window 10 words in length, and a matrix of some 70,000 rows and columns. Each of these rows and columns is labelled with a particular word so as to enable the cells of the matrix to record co-occurences between pairs of words. Co-occurence values for words are assigned by use of the window, with adjacent words being scored 10, adjacent but one 9, and so on. Once this analysis process is complete, rows give the total co-occurence values for words that precede the row label, whilst columns give the same values for words following the column label. For each word, the two 70,000 element vectors are combined to produce a 140,000 element vector representing its similarity to other words, which is HAL's proposed "meaning" of the word.

The HAL vectors correlate strongly with degree of priming in lexical decision tasks (Lund, Burgess and Atchley, 1995), and have been used with mixed success to simulate categorisation (Burgess and Lund, 1997). In this latter study, the vectors for words in four "taxonomic groupings" (geographical locations, cities, animals and body parts) were submitted to multi-dimensional scaling (MDS) to allow a visualisation of the similarities between them. Whilst the MDS showed that HAL could generally separate geographical locations from cities from animals from body parts, clustering each group of these items together, some anomalies did occur, e.g. *eye*, *leg*, and *finger* were closer in the MDS space to *cat* and *dog*, than to *face*, *arm*, or *hand*.

#### *7.4.2.2 LSA*

LSA theory (Landauer and Dumais, 1997) also represents word meanings (or similarities) as vectors, also derived from their co-occurence in text. However its workings are slightly different to those of HAL. In determining LSA vectors, one fist determines a "semantic space," or a set of contexts in which the analysis is to take place (for example, entries in an encyclopaedia can form a set of contexts). The encyclopaedia example uses the first 2000 words of 30,473 entries to form a matrix. Each encyclopaedia entry (context) is assigned a column, and each word a row, and the number of times each word appears in a context is entered into the matrix. As with HAL, a vector is calculated (in this case by logarithmic transformation) which assigns 300-400 values on the same number of dimensions; its LSA "meaning", or similarity vector.

LSA can also be applied to sentence and discourse meanings. Sentences are represented as the average of the vectors of the words they contain, and the coherence between sentences is scored and predicted by calculating - in multidimensional space -a cosine of the angle between the vectors for successive sentences. Landauer and Dumais (1997) claim that this captures the central meanings of passages, and allows similarities between passages to be calculated.

As with HAL, LSA vectors have been shown to have some predictive power: LSA vectors can pick out synonyms with an accuracy that mimics the performance of non-native English speakers (Landauer and Dumais, 1994) and LSA sentence vectors coherence judgements for those sentences (Landauer and Dumais, 1997).

### 7.4.3 Co-Occurence And Semantics

Although LSA and HAL have performed impressively in specific tasks, from the point of view of this thesis, they make unlikely candidates for unified theories of knowledge representation (as Landauer and Dumais, 1997, p 217, suggest).

| *Gentner et al's findings* | Match type | | | |
|---|---|---|---|---|
| | **LS** | **Analogy** | **MA** | **FA** |
| **Soundness** | 4.41 | 4.16 | 2.70 | 2.58 |
| **Similarity** | 4.50 | 4.09 | 3.40 | 2.88 |
| **Proportion Recalled** | .56 | .12 | .53 | .09 |
| **Proportion of keywords recalled** | 1.11 | .17 | .81 | .25 |
| **Quality of reminding** | 1.92 | .44 | 1.64 | .27 |
| | | | | |
| **LSA Cos. (Similarity Score)** | .700 (sd .112) | .527 (sd .157) | .652 (sd .103) | .506 (sd .158) |

**Table 7.1:** Scale of mean ratings of soundness and similarity (compared to the base) and three measures of reminding for the Karla the Hawk story sets, produced by participants in Gentner Ratterman and Forbus (1993), with LSA cosines (similarity ratings) calculated for each story type, also compared to the base.

I have suggested firstly that human concepts have an important structural component, and also that focusing on simple feature lists (and in the end, that is all LSA and HAL analysis amounts to) cannot capture this underlying structure. For example, in experiment 5 (Chapter 6) what was important was not how semantically similar either MOVED_EARLIER or MOVED_LATER was to MOVED_FORWARD (for what it is worth, LSA rates the former most similar), but *which* of the two in particular was to be mapped onto MOVED_FORWARD. And this seemed determined by a particular, rather than a generalised context.

The ability of LSA to make text to text comparisons enabled an experimental approach to be adopted in this respect, since it enabled the calculation of LSA cosines for the various "Karla the Hawk" story sets that participants had rated for similarity in Gentner Ratterman and Forbus's (1993) study (described at length in Chapter 2).

The results of this analysis, and Gentner *et al*'s findings are summarised in Table 7.1[3] (the cosines shown were calculated using a college level psychology text book as the semantic context - reflective of the fact that Gentner *et al*'s participants were all psychology undergraduates - but analyses with other semantic contexts produced the same pattern of results).

As can be seen from this table, LSA cosines are a bad predictor of analogical similarity and participants perceived judgements of inferential soundness in the "Karla the Hawk" stories. Moreover, these results suggest that it is highly unlikely that LSA could model *any* of the empirical findings described in this thesis. Yet if the 'no distinction' hypothesis is correct, these findings are as much a part of conceptual semantics as knowing that "dog" might frequently co-occur with "cat". Chapter 4 showed that there is little hope that the categorisation literature will offer any support for LSA theory here, by proving evidence contra the 'no distinction' hypothesis. Thus it seems, from the evidence presented in this thesis, that any theory of conceptual semantics *ought*, at the very least, to offer some predictive account of the findings of experiments 1, 3, 4 and 5. That LSA, as a theory, cannot do this appears to be directly attributable to the fact that similarity vectors calculated over features in a generalised context are a poor predictor of the effects of structural similarities in a particular context will have. Whereas the evidence reviewed in this thesis appears to show that this latter process is a key ingredient in conceptual semantics.

---

[3] I am grateful to Alex Heneveld and Michael Walker for entering some of this data and calculating some of these cosines.

### 7.4.4 Co-Occurence And Semantic Focus

On the other hand, LSA appears to be a good predictor of recall and reminding (indeed, the one area in which co-occurence models are generally acknowledged to excel is in predicting semantic priming). Furthermore, as a model of retrieval, LSA has been empirically explored to a far greater extent than the retrieval modules of the analogy theories described in Chapter 2, and the apparent success of co-occurance models in tasks such as semantic priming and the identification of synonyms suggests that, from a psychological point of view, there may be something in these models.

The question is what? How can the circle of generalised feature vectors be reconciled with the square of particular structural mappings. An avenue that might be profitably explored here is in generalising and modifying the two-stage model of matching that emerged from the review of analogy in Chapter 2, to produce something like the semantic focus model I started this section with. In such a model, frequency vectors might provide the weak similarities that are sufficient to terminate recursive structure mappings between elements making up the representations that are at the focus of a comparison, but are insufficient to actually support mappings in themselves.

Feature vectors might, therefore, obliquely be able to flesh out Forbus, Gentner, Markman and Ferguson's (1998) speculation about the way acculturation and language learning might equip members of a particular linguistic community with representations of "domains" in terms of canonical sets of dimensions.

An advantage this two-stage theoretical framework possesses is that it abandons the idea that conceptual membership (and hence related notions such as "meaning" and "semantics") is an all or nothing phenomena, a position consistent with both empirical (Chapter 4) and theoretical (Chapter 3) findings in this area, which indicate that questions about category membership are ultimately a matter of context and degree.

If a canonical theory of semantic reconciliation (or re-representation) in comparison mappings is correct (and this is what I interpret Forbus et al's, 1998, proposal to be), then, on our understanding of it, such a theory should predict, and the evidence should show, that an individual's computations of semantic reconciliations should be both stable and consistent.

On the other hand, the theoretical framework of semantic focus (and an alternative reading of Forbus *et al*'s 'acculturation' idea) predicts that such comparisons, reflective of the nature of the semantic reconcilliations that are an integral part of them, should be contingent upon context, and open to revision. Ultimately, conceptual semantics will be dependant upon the relation of individual comparitors to the focus of comparison; upon context and use.

## 7.5 Coming Down To Earth

It should be noted, however, that given the experimental evidence presented in this thesis, this proposal still begs two vitally important questions. Firstly when - at what level - and in what context, the use of generalised information will be sufficient. And secondly, and perhaps most importantly, how is it that the structured representations used in making structural mappings are built up in context?

I have no answers to offer here to these second of these questions.

One possibility relating to the former question is the idea that generalised information has a role to play here as well (if only in terms of attempting to narrow down the potential search space for mappings). However, without better data regarding the second question, relating to the exact nature of just *what* is stored in relation to cognitive concepts, it is difficult to see just how much further speculations such as these can be taken. (And without answers to these questions, all the framework I have described above may amount to is a rather fancy Case-Based Reasoning System.)

In the light of this, the speculative framework offered above should best be seen as an attempt to characterise - to map out - a view of the problem of semantics in analogy (or other comparative mappings) in tractable terms, and not as any final or immutable theoretical position. This has been, to an extent, a negative thesis. I have had more to say regarding what the relationship between analogy and categorisation is not than what it is. These proposals are an attempt to show how this strong negative statement can have positive implications. However, like the mapping process it seeks to describe, this proposal will have to be flexible if it is to be of any use.

## *7.6 Conclusion*

To summarise, the main contributions of the research described in this thesis are:

1. A review of analogical theories, and a specification of the problem of describing analogical semantics.
2. An exegesis of Wittgenstein's "Game" argument (to the author's knowledge, the first such full exegesis of this influential position).
3. A co-ordinated review of the problem of categorisation in cognitive science that places established findings in the framework of Wittgenstein's theoretical analysis.
4. A formulation and exploration of the 'no distinction' hypothesis that holds that there is no principled distinction to be made between literal and non-literal conceptualisation at the cognitive process level.
5. An empirical exploration of the 'no distinction' hypothesis.
6. An empirical analysis of the role of semantic links in analogical mapping.
7. The results of these experiments, that can be used to test future and existing models of analogy and categorisation.
8. A proposal for a future model of conceptualisation.

Thus the primary contribution of this thesis has been an attempt to clarify the relationship between literal and non-literal conceptualisation at the cognitive process level. This is a matter at the heart of cognitive science. I do not claim that this thesis *has* clarified these relationship adequately; rather the claim is that it has at least presented an empirically testable characterisation of this relationship, the 'no distinction' hypothesis, and that this hypothesis can lay the foundation for the scientific investigation of this area.

In addition, the data gathered in the course of this research could be of use to the developers of models of categorisation and analogy. The data from experiments 1 and 3 should provide a good challenge for many classification models, whilst the results of experiments 4 and 5 offer novel and demanding data for analogy models to capture.

## *References*

Baker, GP and Hacker, PMS (1980) *Wittgenstein: understanding and meaning :An analytical commentary on the Philosophical investigations*. Blackwell, Oxford

Bareiss, R and King, J (1989) Similarity assessment in case based reasoning. In Hammond, K (ed.) *Proceedings of the Case-Based Reasoning Workshop*, San Mateo, DARPA, Morgan Kaufman.

Barnes, J (1979) *The Presocratic Philosophers,* Routledge, London

Barresi, J, Robbins, D and Shain, K (1975) Role of distinctive features in the abstraction of related concepts. *Journal Of Experimental Psychology-Human Learning and Memory*, 1, 360-368.

Barsalou, LW (1983) Ad hoc categories Memory and Cognition, 11, 211-227

Barsalou, L (1985) Ideals, central tendency and freuency of instantiation as determinants of graded structure in categories. *Journal Of Experimental Psychology-Learning Memory And Cognition*, 11, 629-654.

Barsalou, L. (1987) The instability of graded structure: implications for the nature of concepts. In Neisser, U. (Ed.) *Concepts and Conceptual Development* Cambridge University Press.

Barsalou, L. (1990) On the indistinguishability of exemplar memory and abstraction in category representation. In: Srull, T.K., Wyers, R.S (eds) *Advances in Social Cognition, Voll. III, Content and Process Specificity in the Effects of Prior Experiences*. Hillsdale, NJ: Erlbaum.

Barsalou, L, Huttenlocher, J and Lamberts, K (1998) Basing categorization on individuals and events, *Cognitive Psychology*, 36, 203-272.

Bechtel, W. & Abrahamsen, A. (1991) *Connectionism and the mind*. Blackwell, Oxford, UK

Berlin, B and Kay, P (1969) *Basic color terms, their iniversalty and evolution,* Berkeley, University of Cakifornia Press.

Black, M (1946) *Language and Philosophy*. Cornell University, Ithaca

Borges, JL (1962) *Labyrinths* Penguin, New York, NY

Boroditsky, L. (1998) Evidence for metaphoric representation: understanding time. In Holyoak, K, Gentner, D and Kokinov, B (eds.) *Advances in Analogy Research,* New Bulgarian University Press, Sofia, Bulgaria.

Boyd, R (1989) What realism implies and what it does not. *Dialectica*, 43, 5 - 29

Boyd, R (1990) Realism, conventionality, and "realism about". In G Boolos (ed.) *Meaning and method: Essays in honour of Hilary Putnam.* Cambridge University Press

Braisby, N, Franks, B and Hampton, J (1996) Essentialism, word use, and concepts. *Cognition.* 59, 247 - 274

Branting, K. (1989) Integrating Generalizations with exemplar-based reasoning. *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society.* Lawrence Earlbaum Assocs., Hillsdale, N.J.

Burgess, C and Lund, K (1997) Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177-210

Burns, B.D. and Holyoak, K.J. (1994) Competing models of analogy: ACME versus Copycat. in *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum Assocs., Hillsdale, N.J.

Burstein, M (1986) Concept formation by incremental analogical reasoning and debugging. In Michalski, Carbonell and Mitchell (eds.) *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufman, Los Altos, CA.

Carrol, J.M. and Mack, R.L., (1985) Metaphor, computing systems and active learning. *International Journal of Man-Machine Studies*, 22, 39-57.

Carey, S (1985) *Conceptual change in childhood.* MIT Press, Cambridge, MA

Clark, H. (1973). Space, time, semantics, and the child. In T Moore (Ed.) *Cognitive Development and the acquisition of language.* Academic Press, New York, NY.

Clark, H and Clark, E (1977) *The psychology of language*. Harcourt Brace Janovich, New York

Clement C.A, and Gentner, D. (1991) Systematicity as a selection constraint in analogical mapping *Cognitive Science*, 15, 89-132.

Clement, J (1988) Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, 12, 563-86.

Cooper, R. & Shallice, T. (1995). Soar and the case for Unified Theories of Cognition *Cognition*, 55 (2), 115-149.

Cooper, R., Fox, J., Farringdon, J. & Shallice, T. (1996) A systematic methodology for cognitive modelling. *Artificial Intelligence*, 85, 3-44.

Cover, T.M & Hart, P.E. (1967) Nearest Neighbour Pattern Classification. *IEEE Transactions on Information Theory,* 13, 21-27.

Davies, I., Corbett, G., Laws, G., McGurk, H., Moss, A. and Smith M.W. (1991) Linguistic basicness and colour processing. *International Journal of Psychology,* 26, 311-327.

Dupré, J. (1981) Biological taxa as natural kinds. *Philosophical Review*: 90, 66-90

Falkenhainer, B., Forbus, K.D. and Gentner, D. (1989) The structure mapping engine: an algorithm and examples. *Artificial Intelligence*, 41, 1-63

Fillmore, C (1971) *Lectures on Deixis*, (reprinted 1997) CSLI Publications, Stanford, Ca

Fish, S (1980) Normal circumstances, literal language, direct speech acts, the ordinary, the everyday, the obvious what goes without saying, and other special cases. In S Fish (ed) *Is there a text in this class?* Harvard University Press

Fodor, JA (1981) The present status of the innateness controversy in J Fodor (ed) *Representations*, MIT Press

Fodor, J & Lepore, E (1996) The red herring and the pet fish: why concepts still can't be prototypes. *Cognition* 58, 253-270

Forbus K, Gentner, D, & Law , K (1995) MAC/FAC: A model of similarity based retrieval *Cognitive Science* 19, 141-205

Forbus, KD, Gentner, D, Markman, AM and Ferguson, RW (1998) Analogy just looks like high level perception: Why a domain general approach to analogical mapping is just right. *Journal of Experimental and Theoretical Artificial Intelligence* 10, 231-358

French, RM (1995) *The Subtlety of Sameness*, MIT Press, Cambridge, Ma.

Gelman, SA and Markman, EE (1986) Categories and induction in young children. *Cognition*, 23, 183-209.

Gentner, D. (1983) Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Gentner, D. (1989) The mechanisms of analogical learning In S. Vosniardou and A. Ortony (eds) *Similarity and analogical reasoning.* Cambridge University Press. pp 200-241

Gentner, D (1997) *Personal Communication.*

Gentner, D.and Landers, R (1985) Analogical reminding: A good match is hard to find. *In Proceedings of the International Conference on Systems, Man and Cybernetics.* Tucson, Arizona.

Gentner, D. and Forbus, K. (1991) MAC/FAC: a model of similarity-based retrieval. *In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society,* Lawrence Earlbaum Assocs., Hillsdale, N.J.

Gentner, D. and Gentner, D. R,. (1983) Flowing water or teeming crowds: mental models of electricity. In *Mental Models.* Lawrence Erlbaum Assocs., Hillside, New Jersey.

Gentner, D Ratterman, MJ & Forbus, K (1993) The roles of similarity in transfer. *Cognitive Psychology* 25, 524-575

Gentner, D and Marman, A B (1997) Structure mapping in analogy and similarity, *American Psychologist*, January 1997, 45-56

Gibbs, RW (1984) Literal meaning and psychological theory. *Cognitive Science* 8:275-304

Gick, M.L. and Holyoak, K.J., (1980) Analogical problem solving. *Cognitive Psychology*, 12, 306-355

Gleitman, LR, Armstrong, SL and Gleitman H (1983) On doubting the concept of "concept" In E Scholnick (ed.) *New trends in conceptual representation,* Lawrence Earlbaum Assoc. Hillsdale, New Jersey

Glucksberg, S. and Keysar, B. (1990) Understanding metaphorical comparisons: Beyond Similarity. *Psychological Review,* 97, 3-18

Goldstone, R.L. (1994) The role of similarity in categorization: providing a groundwork. *Cognition*, 52, 125-157

Goodman, N (1973) Seven strictures on similarity. (in) *Problems and Projects,* Bobbs-Merrill Co., Indianapolis, Indiana

Goodman, N (1976) *Languages of art.* Bobbs-Merrill Co., Indianapolis, Indiana

Gopnik, A and Meltzoff, AN (1997) *Words thoughts and theories,* MIT Press, Cambridge, Ma.

Gould, SJ (1995) *Adams's Navel*, Penguin, London, UK

Hahn, U. & Chater, N. (1997) Concepts and Similarity. in: *Knowledge, Concepts and Categories,* pp. 43-92, K. Lamberts & D. Shanks (eds.), MIT Press.

Hahn, U and Chater, N. (1998) Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable? *Cognition*, 65, 197-203.

Hampton, JA (1979) Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behaviour,* 18, 441-461

Hampton, JA (1981) An investigation into the nature of abstract concepts. *Memory and Cognition*, 9, 149-156.

Harris, R J (1976) Comprehension of metaphors: a test of the two-stage processing model. *Bulletin of the Psychonomic Society*, 8, 312 - 314

Harris R J, Lahey, M and Maraselek, F (1980) Metaphors and images: rating, reporting and remembering. In R Honeck and R Hoffman (Eds.) *Cognition and Figurative Language,* Lawrence Earlbaum Assoc. Hillsdale, New Jersey, 163 - 181

Hoffman, R R and Kemper, S (1987) What could reaction-time studies be telling us about metaphor comprehension? *Metaphor and Symbolic Activity* 2(3), 149 - 186

Hoffman, R R (1995) Monster Analogies. *AI Magazine*, AAAI Press, Fall Issue

Hofstadter, D. (1995) *Fluid Concepts and Creative Analogies.* Basic Books, Harper Collins Publishers, New York, NY

Holyoak, K.J., (1985) The pragmatics of analogical transfer. *Psychology of Learning and Motivation,* 19: 59-87.

Holyoak, K.J. and Koh, K., (1987) *Surface and structural similarity in analogical transfer.* Memory and Cognition, 15, 332-340.

Holyoak, K.J. and Thagard, P. (1989) Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355

Holland, JH, Holyoak, KJ, Nisbett, RE and Thagard, PR (1986). *Induction: Processes of Inference, Learning and Discovery.* MIT Press, Cambridge, Mass.

Homa, D and Vosburgh, R (1976) Category breadth and the abstraction of prototypicalinformation . *Journal Of Experimental Psychology-Human Learning and Memory,* 2, 322-330.

Indurkya, B (1987) Approximate semantic transference: a computational theory of metaphor and analogies. *Cognitive Science*, 11, 445-480

Indurkya, B (1997) Metaphor as change of representation: an artificial intelligence perspective. *Journal of Experimental and Theoretical Artificial Intelligence* 9, 1-36.

Indurkya, B (1992) *Metaphor and cognition*, Kluwer, Dordrecht, Netherlands

Johnson-Laird, PN (1983) *Mental Models.* Cambridge University Press, Cambridge.

Khatchadourian, H (1966) Common names and 'family resemblances' in G Pitcher (ed.) *Wittgenstein: the Philosophical Investigations*, McMillan, London, UK

Hume, D (1739) *A treatise of human nature.* Oxford University Press

Kant, I trans. Meiklejohn, JMD, (1787) *The Critique of Pure Reason,* Dent, London, 1934

Katz, JJ (1972) *Semantic Theory*, Harper & Row, New York.

Keane, M. (1988). Analogical mechanisms. *Artificial Intelligence Review*, 2, 229-251

Keane, M. (1994) Adaptation as a selection constraint on analogical mapping. n *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* Lawrence Earlbaum Assocs., Hillsdale, N.J.

Keane, M.T. (1996). On adaptation in analogy: Tests of pragmatic-importance and adaptability in analogical problemsolving. Quarterly Journal of Experimental Psychology, 49A, 1062-1085.

Keane, M. and Brashaw M (1988) The incremental analogu machine: a computational model of analogy. Technical Report 33, Human Cognition Research Labortory, Open University, UK.

Keane, M., Ledgeway T. & Duff, S. (1994) Constraints on analogical mapping: a comparison of three models, *Cognitive Science* 18, 387-438

Keil, F.C. (1989) *Concepts, kinds and cognitive development.* MIT Press, Cambridge, Mass.

*Kintsch, W (1974)* The representation of meaning in memory. Lawrence Earlbaum Assoc. Hillsdale, New Jersey

Kolers, PA (1972) *Aspects of motion perception*. Pergamon Press, Oxford, UK

Kolers PA and Green M (1984) Color logic of apparent motion.*Perception*, 13, 249-254.

Komatsu, L K (1992) Recent views of conceptual structure. *Psychological Bulletin,* 112(3), 500-526

Kripke, S (1972) N*aming and necessity*. Blackwell, Oxford.

Kruschke, JK (1992) ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.

Labov, W. (1973) The boundaries of words and their meanings. In, Bailey, C-J.N. and Shuy, R.W. (eds) *New ways of analysing variation in English.* Washington, DC: Georgetown University Press. pp 340-373.

Lakoff, G (1987a) *Women, Fire and Dangerous Things: What categories reveal about the mind.* University of Chicago Press, Chicago, Illinois

Lakoff, G. (1987b) Cognitive Models and Prototype Theory. In, Neisser, U. (ed) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge, England: Cambridge University Press.

Lakoff, G and Johnson, M (1980) *Metaphors we live by*. University of Chicago Press, Chicago, Ill.

Lakoff, G and Johnson, M (1999) *Philosophy in the flesh,* Basic Books, Harper Collins Publishers, New York, NY

Lamberts, K. (1996)Exemplar models and prototype effects in similarity-based categorization. *Journal Of Experimental Psychology-Learning Memory And Cognition,* 22, 1503-1507

Landauer, TK and Dumais, ST (1994) Latent semanticanalysis and the measurement of knowledge in R Kaplan and J Burstein (eds.) *Educational service conference on natural language processing techniques and technology in assessment and education*. Educational Testing Service, Princeton, NJ

Landauer, TK and Dumais, ST (1997) A solution to Plato's problem: the latent semanticanalysis theory of acquisition, induction, and representation of knowledge, P*sychological Review*, 104, 211-240

Langley, P. (1996) *Elements of Machine Learning*, San Francisco, CA: Morgan Kaufman.

Law, K, Forbus, K.D. and Gentner, D. (1994) Simulating similarity-based retrieval: a comparison of ARCS and MAC/FAC. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* Lawrence Earlbaum Assocs., Hillsdale, N.J.

Lee, J. R. (1987) *Metalogic and the psychology of reasoning*. Unpublished PhD Thesis, University of Edinburgh

Lister, J, 1997. The Cappilow riot and the composition and behaviour of soccer crowds in late victorian Scotland. *Scottish Footbal Historian*, 66, 8-13.

Lund, K, Burgess, C and Atchley, RA (1995) Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum Assocs., Hillsdale, N.J.

McGee, H (1990) *The Curious Cook*, McMillan, New York, NY

McGinn, C. (1985) *Wittgenstein on meaning : an interpretation and evaluation*: Blackwell, Oxford, UK

McGinn, M (1997) *Routledge philosophy guidebook to Wittgenstein and the Philosophical investigations.* Routledge, London , UK

McGlone, M.S. Conceptual metaphors and figurative language interpreetation: food for thought? *Journal Of Memory and Language*, 35, 544-565

McGlone, M.S. & Harding, J.L. (1998). Back (or forward?) to the future: the role of perspective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 24 , 1211-1223

McTaggart, JE (1908) The unreality of time. *Mind*, 17, 457-474.

McTaggart, JE (1927) T*he nature of existence, Volume 2*, Cambridge University Press, Cambridge, UK

Malt B.C. (1994) Water is not $H_2O$. *Cognitive Psychology* 27:41-70

Malt, BC (1995) Category coherence in cross-cultural perspective. *Cognitive Psychology* 29, 85-148

Malt, BC & Smith, EE (1984) Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behaviour* 23, 250-269

Markman, A.B. and Gentner, D (1993) Structural alignment during similarity comparisons, *Cognitive Psychology*, 25, 431-467

Markman, A.B. and Gentner, D (1997) The effects of alignability on memory, *Psychological Science,* 8 (5) 363-367

Markman, A.B. (1997) Constraints on analogical inference, *Cognitive Science* 21, 373-418.

Medin , D L (1986) Comment on 'Memory storage and retrieval process in category learning.' *Journal of Experimental Psychology: General*, 115, 373 - 381

Medin, DL (1989) Concepts and conceptual structure, *American Psychologist*,44, 1469-1481.

Medin, DL and Ortony, A. (1989) What is psychological essentialism? In S. Vosniardou and A. Ortony (eds) *Similarity and analogical reasoning.* Cambridge University Press. pp 183-196

Medin, DL, Goldstone, R. and Gentner, D. (1993) Respects for similarity. *Psychological Review,* 100, 254-278.

Medin, DL, Goldstone, RL & Markman AB (1995) Comparison and choice: relations between similarity processes and decision processes. *Psychonomic Bulletin* and Review 2(1), 1-19

Melville, H (1851) *Moby Dick; or, The Whale*. Harper Bros. New York.

Mervis, C. B., Catlin, J., and Rosch, E. (1976) Relationships among goodness-of-example, category norms and word frequency. *Bulletin of the Pshychonomic Society* 7, 268-284

Mill, J S (1882) *A system of logic, ratiocinative and induction,* Harper, New York

Miller, G.A., Fellbaum, C., Kegl, J. and Miller, K. (1988) WORDNET: an electronic lexical reference system based upon theories of lexical memory. *Revue Quebecoise Linguistique,* 17, 181-213

Mitchell, M. and Hofstadter, D.R. (1989) The role of computational temperature in a computer model of concepts and analogy-making. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum Assocs., Hillsdale, N.J.

Mitchell, M. (1993) *Analogy making as perception.* MIT Press, Cambridge, Ma.

Murphy, G & Medin, D.L. (1985) The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.

Murphy, GL (1996) On metaphoric representation, *Cognition*: 60, 173-204

Nagel, T (1986) *The view from nowhere*. Cambridge University Press.

Neisser, U, (1987) From direct perception to conceptual structure. In, Neisser, U. (ed) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge, England: Cambridge University Press.

Nosofsky, R.M (1986) Attention, Similarity and the Identification-Categorization Relationship, *Journal of Experimental Psychology: General,* 115, 39-57, Nosofsky, R.M. (1987) Attention and Learning Processes in the Identification-Categorization Relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87-109.

Nosofsky, R.M (1988) Exemplar-based accounts of the relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 700-708.

Nosofsky, R M (1991) Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 3 - 27

Novick, LR (1988) Analogical transfer, problem similarity and expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 14, 510-520

Nosofsky, R.M. (1989) Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception andPsychophysics*, 45, 279-290.

Nosofsky, R.M. (1991) Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental psychology: Human Perception and Performance*, 17, 3-27.

Nosofsky, R.M., Clark, S.E., & Shin, H.J. (1989) Rules and exemplars in categorization, identification and cognition. *Journal of Experimental Psychology: Learning, Memory and Cognition 15*, 282-304.

Ortony, A. (1979) The role of similarity in similies and metaphor. In Ortony, A. (ed) *Metaphor and thought.* Cambridge University Press.

Ortony, A (1980) Some psycholinguistic aspects of metaphor. In R Honeck and R Hoffman (Eds.) *Cognition and figurative language*, Lawrence Earlbaum Assoc. Hillsdale, New Jersey, pp 69 - 83

Ortony, A, Schallert, D L, Reynolds, R E and Antos, S J (1978) Interpreting metaphors and idioms: some effects of context on comprehension. *Journal of Verbal learning and Verbal Behaviour, 17*, 465 - 477

Petrun, C and Belamore, S (1981) Metaphor comprehension and cognitive effort. Paper presented at the *Annual Meeting of The American Psychological Association,* New York.

Pollio, H, Fabrizi, M, sills, A and Smith, M (1982) A note on pauses in spontaneous speech as a test of the derived process theory of metaphor. Unpublished paper, Department of Psychology, University of Tennessee, Knoxville, Tennessee.

Posner, M.I. & Keele, S.W. (1968) On the genesis of abstract ideas. *Journal of Experimental Psychology,* 77, 353-363.

Posner, M.I. & Keele, S.W. (1970) Retention of abstract ideas. *Journal of Experimental Psychology*, 83, 304-308.

Putnam, H (1975a) Is semantics possible? In H Putnam, Philosophical papers: Volume 2. *Mind, language and reality*, Cambridge University Press, Cambridge, U.K. 215 - 271

Putnam, H (1975b) The meaning of "meaning". In H Putnam, Philosophical papers: Volume 2. *Mind, language and reality*, Cambridge University Press, Cambridge, U.K. 139 - 152

Ramscar, M.J.A. and Hahn, U (1998a) What family resemblances are not. in *Proceedings of the 20th Annual Meeting of the Cognitive Science Society, Madison-Wisconsin*, LEA, NJ, pp 865-870.

Ramscar, M.J.A. and Hahn, U. (1998b) Wittgenstein and the ontological status of analogy. In Holyoak, K, Gentner, D and Kokinov, B (eds.) *Advances in Analogy Research,* New Bulgarian University Press, Sofia, Bulgaria, pp 390-402

Ratterman, M.J. and Gentner, D. (1987) Analogy and similarity: determinants of accessibility and inferential soundness. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society,* Lawrence Earlbaum Assocs., Hillsdale, N.J.

Récanati, F (1995) The alleged priority of literal interpretation Cognitive Science 19, 207-273

Reed, S K (1972) Pattern recognition and categorisation. *Cognitive Psychology*, **3**, 382 - 407

Rips, L.J. (1975) Inductive judgements about natural categories *Journal of Verbal Learning and Verbal Behaviour* 14,665-681

Rips, L.J. (1989) Similarity, typicality and Categorization. In S. Vosniardou and A. Ortony (eds) *Similarity and analogical reasoning*. Cambridge University Press.

Rosch, E. and Mervis, CB (1975) Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* **3**:382-439

Rosch, E, Simpson, C. & Miller, RS (1976) Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance* 2, 491-562.

Rosch (1975) Cognitive reference points. *Cognitive Psychology*, 7, 532-547.

Rosch (1978) *Cognition and Categorisation*. Lawrence Earlbaum Assoc., New Jersey.

Ross, B.H. (1984) Remindings and their affects in learning a cognitive skill. *Cognitive Psychology,* 16, 371-416

Ross, B.H. (1987) This is like that: the use of earlier problems and the separation of similarity of effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 629-639

Roth, EM and Shoben, EJ (1983) The effect of context on the structure of categories. *Cognitive Psychology*, 15, 346-278

Rumelhart, D.E (1979)Some problems with literal meanings. In A. Ortony (ed) *Metaphor and Thought.* Cambridge University Press.

Rumelhart, D.E. and Abrahamson, A.A. (1973) A model for analogical reasoning. *Cognitive Psychology*, 5, 1-28

Rumelhart, D.E. and Norman, D.A. (1981) Analogical processes in learning. In Anderson, J.R. (ed), *Cognitive skills and their acquisition*, Lawrence Earlbaum Assocs., Hillsdale, N.J.

Rundle, B (1996) Wittgenstein and contemporary philosophy of language. Blackwell, Oxford, UK

Schumacher, R.M. and Gentner, D. (1988) Transfer of training as analogical mapping. *IEE Transactions on Systems, Man and Cybernetics*, 18, 592-600

Schyns, P.G. & Rodet, L. (1997) Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory and Cognition,* Vol. 23, 681-696.

Schyns, P.G., Goldstone, R.L., & Thibaut, J. (in press) The development of features in object concepts. *Behavioural and Brain Sciences*

Shin, H.J. and Nosofsky, R.M. (1992) Similarity-Scaling Studies of Dot-Pattern Classification and Recognition. *Journal of Experimental Psychology: General,* 121, 278-304.

Skorstad, J., Falkenhainer, B. and Gentner, D. (1987) Analogical processing: a simulation and empirical corroboration. In Proc. *Meeting of the American Association of Artificial Intelligence,* Morgan Kaufman, Los Altos, Ca.

Sloman, S., Malt, B., Shi, M., Gennari, S. and Wang, Y. (1997a) The relationship of similarity to naming: Chinese vs. American conceptions of bottles and jars. in *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum Associates, New Jersey, pp 697-702

Sloman, S., Malt, B., Shi, M., Gennari, S. and Wang, Y. (1997a) Are bottles similar to oneanother? Sorting and naming by Chinese, Argentinians and Americans. in *Proceedings of SimCat 97*, Department of Artificial Intelligence Conference Proceedings, University of Edinburgh, Scotland, pp 233-240

Small, S.L. (1997) Semantic category imprecision: A connectionist study of the boundaries of word meanings. *Brain and Language,* 57, 181-194.

Smith, E.E. & Medin, D.L. (1981) *Categories and Concepts.* Cambridge, MA: Harvard University Press.

Smith, E.E, Shoben, E.J., and Rips, L.J. (1974) Structure and process in semanticmemory: a feature model for semantic decisions *Psychological Review* 81, 214-241

Spellman, B. & Holyoak K. (1992) If Saddam is Hitler, then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology,* 62, 913-33

Spellman, B. & Holyoak K. (1996) Pragmatics in analogical mapping, *Cognitive Psychology,* 31, 307-346.

Sternberg, R.J. (1977) Component processes in analogical reasoning. *Psychological Review*, 84, 353-378

Taylor, J.R. (1995) *Linguistic categorization: Prototypes in linguistic theory*. 2nd edition. Oxford University Press.

Thagard, P., Holyoak, K.J., Nelson, G. and Gochfield, D. (1990) Analog retrieval by constraint satisfaction. *Artificial Intelligence*, 46, 259-310

Traugott, E. C. (1978). On the expression of spatiotemporal relations in language. In J. H. Greenberg (Ed.), *Universals of human language: Vol. 3. Word structure*, Stanford University Press, Stanford California, pp.369-400

Wasserman, E. R. (1960) *Pope's epistle to Bathurst: a critical reading*. John Hopkins Press, Baltimore, MA

Wharton, C.M., Holyoak, K.J., Downing, P. E., Lange, T. E., Wickems, T. D. and Melz, E. R. (1994) Below the surface, Analogical similarity and retieval competition inreminding. *Cognitive Psychology*, 26, 64-101

White, R.H (1987) A critical comparison of models of analogical reasoning and their application to learning. DAI Discussion Document 38, Dept of Artificial Intelligence, University of Edinburgh.

White, R.H. and Brna, P (1990) The analogical constraint mapping engine - an investigation. Unpublished working paper.

Williams, R (1976) *Keywords. A vocabulary of culture and society*. Collins-Fontana, Glasgow

Winston, P.H (1980) Learning and reasoning by analogy. *Communication of the ACM*, 23, 689-703

Wittgenstein, L trans. Ogden, C.K. (1922) *Tractatus logico-philosophicus*, Routledge & Kegan Paul, London

Wittgenstein, L trans. Anscombe, E (1953). *Philosophical Investigations* Blackwell, Oxford.

# Appendix A: Materials Used In Experiments 1 and 2

The Karla the Hawk story sets from Gentner, Ratterman and Forbus (1993).

SET1

**BASE Story:**

William was a patient in a psychiatric hospital who was confined indoors almost all the time. He could never pass the monthly room inspections so he hated them. He spent most of his time daydreaming about food. A few days before the April inspection William's room was still a mess since he had done nothing but daydream. To provide William with an incentive, the nurse promised him some gingerbread from the cookie shop if he scrubbed his room and put it in order once and for all. William was overjoyed. But there was no longer enough time for him to put it in order. As a result, he did not pass the inspection and did not get any gingerbread.

**LS:**

Kevin was shut-in at a mental institution. He always did poorly at the weekly bedroom checks so he disliked them. About the only thing Kevin liked to do was eat.

Not long before the last check of the year, Kevin's bedroom was still a disaster because he had spent all his time daydreaming. To motivate him, the matron promised Kevin a cake from the bakery if he cleaned and organised his bedroom.

This made Kevin extremely happy. But he was too far behind in his cleaning to get his bedroom ready in time. Consequently, he didn't get a cake at all.

**MAAO**

There once was a boy named William who thought about food a lot. He had a particular liking for gingerbread. One of the happiest times of his childhood was spent on holiday with his aunt. He got to sit indoors all day reading comics. His aunt worked as a nurse in a psychiatric hospital.

**TA:**

Karen always did poorly in high school so she despised it. But she loved vacations. She spent most of her time dreaming about going to Hawaii. Not long before the end of her fourth year Karen was not doing at all well in her classes because she had spent all her time daydreaming. To motivate her, Karen's father promised her that if she did well enough during the next few week to graduate from high school he would pay for her trip to Hawaii. This made Karen extremely happy. But she was too far behind in her classes. Consequently she failed too many and did not go to Hawaii.

**FA:**

Karen disliked high school so she always had trouble getting passing grades in her classes. Vacations were her favourite part of the school year. Not long before the end of her fourth year of school, Karen was not doing well at all because she despised school. To motivate her, Karen's father promised her a trip to Hawaii if she would just pass enough of her classes to graduate. This made Karen extremely happy. During the last few weeks of school she spent most of her time dreaming about Hawaii and preparing for her trip there. Consequently, she failed to graduate and she did not get to go.

**MA:**

Kevin was shut-in at a mental institution. He disliked cleaning his room so he always failed the weekly bedroom checks. About the only thing Kevin liked to do was eat.

Not long before the last check of the year, Kevin's bedroom was still a disaster because he despised cleaning. To motivate him, the matron promised Kevin a cake from the bakery if he cleaned and organised his bedroom.

This made Kevin extremely happy. During the next few days he spent most of his time dreaming about the bakery and deciding what kind of cake to get. Consequently, he didn't clean his bedroom and did not get a cake after all.

SET 2

**BASE Story:**

Mr Newton was the manager of a company that made razors. One year, an inventor in his company perfected a metal that would stay sharper than any previous razor.

However, not long after that, Mr Newton was using one of these new razors and he cut his neck rather severely. As a result of this incident the inventor recommended a safety feature that would protect people from injuring themselves.

After what had happened, Mr Newton considered the safety feature absolutely necessary. The incident had alarmed him so much that he began to produce the safety razor right away.

**LS:**

Mr Boyce was director of manufacturing shaving knives for Gilette Company. In 1980 an engineer in his company perfected a steel alloy that would keep a finer edge than any previous blade. But later, when Mr Boyce was using one of these new knives he slashed his throat pretty seriously. Because of this accident, the engineer recommended a new protective part that would prevent people from hurting themselves with the blade.

After this, Mr Boyce considered the protective part absolutely necessary. The accident frightened him so much that he had the protective part developed immediately.

**TA:**

Ms Boyce was in charge of installing heat lamps sold by the Smith Tan Company. In 1980, a supplier for the company perfected a filament that could produce more intense heat than any previous filament. But later, when Ms Boyce was using one of these new heat lamps herself she was seriously burned. Because of this accident, the supplier recommended a warning label that would protect people from hurting themselves with the new lamp. After this, Ms Boyce considered the labels absolutely necessary. The accident frightened her so much that she had the labels printed immediately.

**FA:**

Ms Boyce was in charge of installing sun lamps sold by Smith Tan Company. In 1980, a supplier for the company perfected a filament that could produce more intense heat than any previous filament. Because it was so hot, the supplier recommended a warning label that would protect people from hurting themselves with the new lamp. But Ms Boyce considered the labels unnecessary. It would be too much trouble to make them. But later, when she was using one of these new heat lamps herself, she was seriously burned. After this accident she was so frightened that she had the warning labels printed immediately.

**MA:**

Mr Boyce was director of manufacturing shaving knives for Gilette Company. In 1980, an engineer in his company perfected a steel alloy that would keep a finer edge than any previous blade. Because it was so sharp, the engineer also recommended a new protective part that would prevent people from hurting themselves with the blade.

But Mr Boyce considered the protective part unnecessary. It would be too expensive to incorporate.

But later, when he was using one of these new blades, he slashed his own throat pretty seriously. After this accident he was so frightened that he had the protective part developed immediately.

**MAAO**

There was once an engineer who loved his job. He thought all day about complex engineering problems. He was proud that he invented a steel alloy for improving shaving knives. He was working in his garage when the idea came to him. His friend, the director of manufacturing, was with him at the time.

**SET 3**

**BASE Story**

Joseph was a millionaire who hired a chauffeur to drive his Rolls-Royce. He used to brag to his wife that he would never be late for his conferences since he had hired a chauffeur.

One morning when he was in a great hurry, he went to find the chauffeur. But he was asleep. He thought his services would not be needed that day.

Thus Joseph was very late for his conference after all. To make sure this would not happen again, Joseph hired a second chauffeur.

Removed:

This turned out to be very expensive because the second chauffeur wanted a fancy uniform as well as a high salary.

**LS:**

Alexander was a wealthy man who employed a driver for his limousine. He liked to boast to his spouse that with his driver he would always be on time for his meetings.

One day when he was in a rush, he went to find the driver. But he was taking a nap. The driver thought it was his day off. Thus Alexander ended up missing his meeting. But to make sure he would not be late again, Alexander hired a second driver.

**TA:**

Alexander was a man who lived with his wife in his house a long time ago. He liked to boast to a friend that with his wife at home we would always eat well. One day, when he was very hungry, Alexander went home to his wife. But she thought Alexander would be eating someplace else so she had only prepared enough for herself and their baby. Thus Alexander went without dinner. But to make sure he would not go hungry again, Alexander married a second wife.

**FA:**

Many years ago there was a man named Alexander who had two wives living in two houses. He liked to boast to a friend that with his two wives he would always eat well. One day when he was very hungry, Alexander went to his first wife. But she had only prepared enough for herself and their baby. She thought Alexander would eat with his other wife that day. Alexander then went to his second wife, but found she had already finished. She saw him go to the other house so she thought he was eating there. Thus Alexander went without any dinner at all.

**MA:**

Alexander was a wealthy man who employed two drivers for his two limousines. He liked to boast to his spouse that with his two drivers he would always be on time for his meetings.

One day, when he was in a rush, Alexander went to the first driver. But he was taking a nap. The driver thought it was his day off.

Then Alexander called for the other driver, but he had already left. He had seen Alexander with the first driver so he thought he had the day off.

Thus, Alexander ended up missing his meeting.

**MAAO**

Clive was a chauffeur. He was proud of the fact that he always got his boss to meetings on time. Even when his boss was late leaving, Clive made sure that they were not late for the conference. His boss's car was a Rolls Royce. Clive's own car was a Ford. It was an automatic, whilst his boss's car had manual gears.

**SET 4**

**Story:**

Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so grateful that he pledged never to shoot at a hawk again.

**LS:**

Once there was an eagle named Zerdia who nested on a rocky cliff. One day she saw a sportsman coming with a crossbow and some bolts that had no feathers. The sportsman attacked but the bolts missed. Zerdia realised that the sportsman wanted her tailfeathers so she flew down and donated a few of her tailfeathers to the sportsman. The sportsman was pleased. He promised never to attack eagles again.

**TA:**

Once there was a small country called Zerdia that learned to make the world's smartest computer.

One day Zerdia was attacked by its warlike neighbour, Gagrach. But the missiles were badly aimed and the attack failed. The Zerdian government realised that Gagrach wanted Zerdian computers so it offered to sell some of its computers to the country. The government of Gagrach was very pleased. It promised never to attack Zerdia again.

FA:

Once there was a small country called Zerdia that learned to make the world's smartest computer. Zerdia sold one of its supercomputers to its neighbour, Gagrach, so Gagrach would promise never to attack Zerdia.

But one day Zerdia was overwhelmed by a surprise attack from Gagrach. As it capitulated, the crippled government of Zerdia realised that the attacker's missiles had been guided by Zerdian supercomputers.

**MA:**

Once there was an eagle named Zerdia who donated a few of her tailfeathers to a sportsman so he would promise never to attack eagles.

One day Zerdia was nesting high on a rocky cliff when she saw the sportsman coming with a crossbow. Zerdia flew down to meet the man, but he attacked and felled her with a single bolt. As she fluttered to the ground Zerdia realised that the bolt had her own tailfeathers on it.

**MAAO**

There once was a sportsman who loved to hunt. He liked to have the animals he caught stuffed and mounted. His pride and joy was an eagle he had killed with just a crossbow and a bolt. He had been hiding in the top of an elm tree when he shot her.

**SET 5**

**BASE Story:**

Peter was a young man from Iceland who moved to Florida one summer. He was very self-conscious about his pale skin, though. Peter thought that people would not accept him unless he had a dark tan like everyone else.

So, Peter spent a whole day in the sun, trying to get a tan. H e didn't understand how dangerous the sun could be. Consequently, by evening, he had to go to the hospital. He had second degree burns over most of his body.

**LS:**

Alfred was a young man from Sweden who was vacationing in Hawaii. He was annoyed at having to remain pale. He wanted to impress people with his dark tan.

So one day, Alfred spent the whole afternoon in the sun. He didn't realise what a risk being in the sun could be. As a result, by nightfall he had a terrible sunburn and needed medication. Alfred decided that Hawaii was not his style and decided to fly back to Sweden.

**TA:**

Quiggly Company was a small company that manufactured electrical components. One year, Quiggly Company began to get into the integrated circuit market. However, Quiggly Company felt that to be competitive it would have to have really sophisticated space factories, as did all the other manufacturers of integrated circuits. Thus, Quiggly put all its assets into having a space factory. The company just didn't realise what a financial risk space manufacturing was. As a result, within six months, Quiggly Company had spent much more than they could afford and their stocks plunged to an all time low. Quiggly Company decided to putt out of integrated circuits altogether and went back to making conventional electrical components.

**FA:**

Quiggly Company was a small company that manufactured conventional electrical components. They had just started producing integrated circuits and decided that the real action in the integrated circuit market was in space manufacturing. But the market analysts gave them a stern warning. They said that space factories were too expensive and that the market was too competitive. If Quiggly went into space manufacturing, the analysts predicted they would lose everything and their stock would plunge.
Quiggly Company was so annoyed at having to play it safe when everyone else had a chance to make a killing that they decided to go back to manufacturing conventional electrical components.

**MA:**

Alfred was a man from Sweden who was vacationing in Hawaii. He wanted to impress his friends with a dark tan, so one day he went out to spend the afternoon in the sun. But a doctor acquaintance chanced to pass by and gave Alfred a stern warning. If Alfred stayed in the sun, by nightfall he would have a terrible sunburn and would need medication.

Alfred was so annoyed at having to remain so pale when everyone else was tan that he decided to go back to Sweden.

**MAAO**

There once was a Norwegian called Brunhilde who liked beach holidays. She loved to tan her white skin a nice shade of brown. She went to Florida one year on a package holiday. The holiday was expensive but she didn't mind. She travelled from Norway to America on a jumbo jet.

**SET 6**

**BASE Story:**

Julius was a mule who discovered several pears sitting in a window sill. He thought to himself, "These pears seem to be rotten. Perhaps I'll get some and find out if my prediction is correct."

However, the pears were too high up for Julius. And because he was hungry he felt too weak to jump up to them. Naturally, this was rather disappointing.

**LS:**

A pony named Sidney found some apples resting on a beam in the barn. "Those apples really look wormy," Sidney thought to himself. "I think I'll try to get some so I can see if my hunch is right."

But the apples were out of reach. Needless to say, he could not climb up the ladder to get them. Naturally Sidney became somewhat disappointed.

**TA:**

A girl named Cindy found some records she was curious about at a record store. "These records look really bad," Cindy thought to herself. "I think I'll buy them so I can see if my hunch is right."

But, the records were too expensive for her. After thinking it over, she saw there was no way she could afford them. Needless to say, she became somewhat disappointed.

**FA:**

A girl named Cindy found some records she wanted at a record store. She really wanted them but, needless to say, they were too expensive for her.

After brooding for a while, Cindy got over her disappointment.

"These records are probably boring," she thought. "So even if I could afford them I wouldn't listen to them."

**MA:**

A pony named Sidney found some apples resting on a beam in the barn. He was starving, but they were out of reach. Needless to say, he could not climb up the ladder to get them.

After brooding for a while, Sidney got over his disappointment.

"These apples are probably all wormy," he thought. "So even if I could get one I would not want to eat it."

**MAAO**

Olivia Aristotle loved pears. She always bought them from Mr Cucumber, her grocer. Usually she left her pears on the window sill. Sometimes she gave a pear to her mule. The mule lived in her garden.

**SET 7**

**BASE Story:**

Percy the mockingbird spent the whole warm season chirping and twittering. When it began to get colder Percy visited a squirrel and sang a song for her, expecting to get some of the squirrel's sunflower seeds in return. However, the squirrel was very disappointed in him.

"You are a terrible singer!" she yelled. "I'm not giving you any of my wheat."

A tear rolled down Percy's cheek, and he vowed to give up singing for good.

**FA:**

Sam travelled all over the world buying beautiful things. When he ran out of money he paid a visit to his mother. However, she was not at all pleased with him.

"While I have been hard at work you have been wasting your time," she said. Sam gave her a gift he bought in Tibet, hoping she would give him a loan in return. But she was still not pleased. "I will not give you any of my hard-earned money!" she exclaimed.

**TA:**

Sam travelled all over the world buying beautiful things. When he ran out of money he paid a visit to his mother and gave her a gift he bought while in Tibet, Hoping she would give him a loan in return. However, his mother was not at all pleased. "You don't deserve any money of mine!" she exclaimed. "This is a piece of junk!"

**LS:**

A magpie named Sam sang all summer. When winter came he paid a visit to a chipmunk and performed a ballad for her hoping she would give him some nuts in return. However the chipmunk was not at all pleased.

"You don't deserve any nuts of mine!" she exclaimed. "Your song was terrible."

**MA:**

A magpie named Sam sang all summer. When winter came he paid a visit to a chipmunk. However, the chipmunk was not at all pleased with Sam.

"You have wasted the summer while I have been hard at work!" she said. Sam performed a ballad for her hoping she would give him some nuts in return. But she was still not pleased. "I will not give you any of my nuts!" she exclaimed.

**MAAO**

There was once a mockingbird who loved singing. She liked to stand up on the top of trees when she sang. One day Hank the squirrel offered the mockingbird a recording contract. He had been eating nuts when he first heard her sing.

**SET 8**

**Story:**

A dog named Leonard was trotting along a dock holding a steak in his teeth. However, as he trotted he dropped it inadvertently and it disappeared into the lake. Thus, he was left without his dinner.

Then Leonard saw another dog with a steak. Because he had lost his own steak he tried to take it. Unluckily, the other dog was much bigger than he was.

**FA:**

Charlie was walking through a hotel with an expensive pocket watch in his hand. Then he noticed another man with an expensive watch.

Not realizing it was just his reflection, and since he was also quite greedy, Charlie reached out to take it. Unfortunately, as he opened his hand he dropped his own watch and both watches shattered on the floor. Consequently he was left with no watch at all.

**TA:**

Charlie was walking through a hotel with an expensive pocket watch in his hand. As he walked he accidentally dropped it and it shattered on the floor. Consequently, he was left without a watch.

Then Charlie saw another man with an expensive watch. Since he had broken his own watch he reached out to take it. Unfortunately, the man was a policeman.

**LS:**

Charlie the puppy was running out on a pier with a pork chop in his mouth. As he ran he accidentally dropped it and it vanished into the bay. Thus, he was left with nothing to eat.

Then Charlie saw another puppy with a pork chop in his mouth. Since he had lost his own pork chop he reached out to take it. Unfortunately, the other puppy was much larger than he was.

**MA:**

Charlie the puppy was running out on a pier with a pork chop in his mouth. Then he looked down and saw another puppy with a pork chop in his mouth.

Not realizing it was just his reflection, and since he was also quite hungry, Charlie reached down to take it. Unfortunately, as he opened his mouth he dropped his piece and both pieces vanished into the bay. Consequently, he was left with no meat at all.

**MAAO**

There was once a man called Ludwig who had a dog. On payday he always bought his dog a steak. He lived in a house by a lake. Sometimes he went fishing with his friend Hal, who also had a dog. Hal's dog was bigger than Ludwig's dog.

**SET 9**

**BASE Story:**

Boris and Ivan thought well of one another's skill in business and resolved to open up a store together. As ill luck would have it, Ivan was quite absent-minded and he threw out a large amount of cash. This annoyed Boris who therefore demanded that Ivan have nothing to do with the monetary matters of their new store.

**FA:**

John and Christine loved each other and decided to be married. Unfortunately, John discovered that Christine was a very reckless driver. So he insisted that Christine never drive his new car. This upset Christine so much that she dented John's new car hoping that John would be hurt.

**TA:**

John and Christian loved each other and decided to be married. Unfortunately, Christian was so reckless that she accidentally dented John's new car. This upset John, so he insisted that she never drive his car again.

**LS:**

John and Christian respected each other's technical judgment and decided to form a company. Unfortunately, Christian was so forgetful that he put all their money in the trash. This upset John, so he insisted that Christian never deal with the proposed company's finances.

**MA:**

John and Christian respected each other's technical judgment and decided to form a company. Unfortunately, John learned that Christian was notoriously forgetful, so he insisted that Christian never deal with the proposed company's finances. This upset Christian so much that he put all their money in the trash to get back at John.

**MAAO**

There was once two friends who decided to become business partners. They saw an opening in the recycling business. The business was a great success. They used to say to one another "There's good money in old garbage". They enjoyed their work a lot.

**SET 10**

**Story:**

Two small countries, Bolon and Salam, were adjacent to a large, warlike country called Mayonia, which had a huge fleet of destroyers and military jets.

Bolon decided to make the best of the situation by taking over Salam. Salam started looking for aid from other strong countries but soon Bolon succeeded in taking over a coastal city.

Then victorious Bolon proposed to make a treaty with its warlike neighbour Mayonia. Bolon proposed to give Mayonia control over Salam in exchange for a guarantee that Bolon would remain independent.

Mayonia responded by overrunning both Bolon and Salam. Bolon was so busy maintaining control of Salam, it could do nothing to stop Mayonia. Thereupon, Mayonia installed puppet governments in both Bolon and Salam, and took over the newspapers and radio stations.

**FA:**

Two seventh graders, Lincoln and Moreland, were walking to school together when they met Chad. Chad was a mean high school boy who was known to rob younger kids.

Moreland was scared but he decided to make the best of the situation by offering to help Chad rob Lincoln in exchange for being left alone himself.

But Lincoln overheard and started running to find a policeman. Immediately, Moreland took off after him and eventually succeeded in catching him.

When Chad arrived he robbed both of them. And Moreland was so exhausted, he could do nothing to prevent it.

**TA:**

Two sixth-grade boys, Lincoln and Moreland, were walking to school together when they met Chad. Chad was a mean high school boy who was known to rob younger kids.

Moreland was scared but he decided to make the best of the situation by overpowering Lincoln. Immediately, Lincoln started running to find a policeman but Moreland ran after Lincoln and eventually succeeded in catching him.

When Chad arrived Moreland offered to let him rob Lincoln in exchange for being left alone himself. But Chad robbed both of them. Moreland was so busy holding Lincoln down, he could do nothing to prevent it.

**LS:**

Two weak nations, Lincoln and Moreland, bordered a third nation known as Chad. Chad was aggressive and very powerful. Moreland decided to make the best of the situation by overpowering Lincoln. Lincoln started to seek protection from other powerful nations but Moreland eventually succeeded in conquering it.

Then triumphant Moreland offered to make a pact with its aggressive neighbour Chad. Moreland offered to give Chad control over Lincoln in exchange for being left alone itself.

When Chad got into the action it invaded both Lincoln and Moreland. And Moreland was so drained from battle, it could do nothing to prevent it.

**MA:**

Two weak nations, Lincoln and Moreland, bordered a third nation known as Chad. Chad was aggressive and very powerful. Moreland decided to make the best of the situation by offering to help conquer Lincoln in exchange for being left alone itself.

But Lincoln somehow received word of the agreement and began to seek protection from other powerful nations. Immediately, Moreland invaded Lincoln and eventually succeeded in conquering it.

When Chad got into the action it invaded both Lincoln and Moreland. And Moreland was so drained from battle, it could do nothing to prevent it.

**MAAO**

Once there was a very warlike small country. It liked to invade other countries and subjugate their people. Its greatest moment in history was the conquest of a small neighbouring island. The main assualt had taken place at dawn.

**SET 11**

**BASE Story:**

A farmer named Beedle had a choice apricot tree that he valued above all the rest in his garden. Every year Beedle gave presents of fruit from his tree to everyone he knew. But one year the tree began to wither and there was no more of the wonderful fruit. Everyone was very unhappy. The judge had enjoyed the apricots from the tree very much - so much that he resolved to take the tree to his own garden to take care of it. He paid Beedle very well and had the tree moved to his own grounds. Fortunately, as soon as it was transplanted, the tree bore delicious apricots again. That very year, the fruit won first prize in the county fair.

**FA:**

Mr Gerson had several pets but he valued the canary above all the rest. Occasionally, he invited friends over to hear the canary sing.

A wealthy businessman liked the canary's songs so much that one autumn he decided he wanted the bird in his house in the city where he could enjoy it every day. He paid Mr Gerson very well and had the canary moved to his house.

However, upon being moved the canary became fatally ill and there were no more of its beautiful songs. Everyone was extremely disappointed.

**TA:**

Mr Gerson had several pets but he valued the canary above all the rest. Occasionally, he invited friends over to hear the canary sing.

However, one autumn the canary became very ill and there were no more of its beautiful songs. Everyone was extremely disappointed.

A veterinarian friend had liked the canary's songs so much that he decided he would buy the canary and put it in his office where he could look after it. He paid Mr Gerson very well. Upon being moved, the canary sang beautiful songs again.

**LS:**

A grower, Mr Gerson, had many grapevines in his vineyard but there was one that he valued above the rest. Occasionally, Mr Gerson presented grapes from this vine as gifts to his friends.

However, one autumn the vine began to dry up and there were no more of the marvellous grapes. Everyone was extremely disappointed.

The mayor liked the grapes from this vine so much that he decided he would buy the vine and put it in his own yard where he could look after it. He paid Mr Gerson very well and had the vine removed to his estate. Upon being moved, the vine produced marvellous grapes again.

**MA:**

A grower, Mr Gerson, had many grapevines in his vineyard but there was one that he valued above all the rest. Occasionally, Mr Gerson presented grapes from this vine as gifts to his friends.

The mayor liked the grapes from this vine so well that one autumn he decided he wanted the vine in his own yard where he could enjoy them every day. He paid Mr Gerson very well and had the vine moved to his estate.

However, upon being moved, the vine dried up and there were no more of the marvellous grapes. Everyone was extremely disappointed.

**MAAO**

An apricot farmer once had a favourite apricot tree. He gave his friends apricots from the tree. Eventually, the tree finally withered and died. The farmer made a stool from the tree wood. That year judge's tree won the prize at the country fair.

**SET 12**

**BASE Story:**

Morris was a simple-minded prisoner who loved to play the harmonica at night. Finally, his cellmate got tired of this and demanded that he shut up. He should always play during the day, the cellmate insisted.

The next night, Morris escaped from prison and hid in a boxcar. But then Morris remembered that his cellmate said he should play the harmonica every day, so he started playing. Immediately, Morris was nabbed by the cops. Poor, simple-minded Morris wished he hadn't paid any attention to his cellmate. Back in his cell, he devoted himself to covering the walls with obscenities.

**FA:**

A little girl named Jenny wished she could run away from home. She loved to play with her dog every day, but only in her neighbour's yard. Eventually, her neighbour got fed-up and told her to always play in her own yard.

But Jenny refused to listen because she thought it was bad luck to play in her own yard. The last time she did it her mother caught her right away and made her go to work.

**TA:**

A little girl named Jenny loved to play with her dog in her neighbour's yard. Eventually, the neighbour got fed-up and told her she should always play with her dog in her own yard.

A few days later, Jenny snuck out of the house without doing her chores and hid at the playground. But after a while she went home to play with her dog because she thought her neighbour said she should every day. Right away her mother caught her and the little girl wished she hadn't listened to her neighbour.

**LS:**

A convict named Denny loved to play the kazoo, but he would only play it long after dark. Eventually the convict in the next cell got fed-up and told him he should always play during the day time.

A few days later, Denny snuck out of jail and hid in a railroad car. But after a while he began to play the kazoo because he thought the convict in the next cell had said he should play every day. Right away the police caught him and Denny wished he hadn't listened to the convict in the next cell.

**MA:**

A convict named Denny wished he could break out of jail. He loved to play the kazoo in jail every day but he would only play it long after dark. Eventually, the convict in the next cell got fed-up and told him to stop and always play during the daytime.

But Denny refused to listen because he thought it was bad luck to play during the day. The last time he had done that guards had caught him right away and thrown him into solitary confinement.

**MAAO**

Once there was a prisoner who played the harmonica. He played "Old swanne river" very badly. One day his cell mate escaped. He was recaptured in a boxcar. The cellmate's name was Crusher Spiggins.

**SET 13**

**Story:**

King Otto was a warmonger who was obsessed with conquest and riches. One year he decided to rebuild all the bridges of the kingdom to make them strong enough for his armies.

Otto's counsellor, the chamberlain, complained that he was spending too much money on the bridges and not enough on actual military campaigns. Because of all the construction too many of the bridges were out of commission being worked on. Otto decided that the chamberlain was right and he abandoned the bridge project to mount a full-scale invasion of a neighbouring country.

Then one day when King Otto was travelling through his kingdom his carriage crashed through an unfinished bridge. He was severely injured but the mishap made it clear to him that he needed to balance his military and domestic expenditures. Also, he ordered the chamberlain to get him a new crown because the old one had gotten smashed in the crash.

**FA:**

Cornelius was an astronomer who thought about nothing but astronomy. He was obsessed with stars and galaxies.

His wife Agatha complained that he was spending too much time on astronomy and not enough on maintaining his own house. As a result, the roof was gradually leaking more and more and the house was becoming almost unlivable. But Cornelius would not listen.

Then, one evening when Cornelius was walking through his house, a rotten floor collapsed under his weight. Cornelius was severely injured but the accident showed him that Agatha had been right all along. He decided to try to balance his time between astronomy and maintaining his home.

**LS:**

Emperor Cornelius thought about nothing but war. He was obsessed with victory and treasure. One year he decided to renovate all the roads of the empire to make them safe enough for his armed forces.

But before the roads were finished, the emperor's closest adviser, the prime minister, complained that he was spending too much of the royal treasury on the roads and not enough on actual warfare. As a result, the roads were constantly under construction and were seldom usable. Cornelius decided that the prime minister was right and he went back to thinking about nothing but war.

Then one evening when Cornelius was riding in the countryside, an unmaintained overpass collapsed under his horse-drawn coach. He was severely injured but the accident showed him that he should try to balance his spending between war and maintaining his empire.

**MA:**

Emperor Cornelius thought about nothing but war. He was obsessed with victory and treasure.

The emperor's closest adviser, the prime minister, complained that he was spending too much of the royal treasury on war and not enough on maintaining his own empire. As a result, the roads of the empire were being neglected and they were gradually beginning to deteriorate. But Emperor Cornelius didn't listen.

Then, one evening when Cornelius was riding in the countryside, an unmaintained overpass collapsed under his horse-drawn coach. He was severely injured but the accident showed him that the prime minister had been right all along. Cornelius decided to try to balance his spending between war and maintaining his empire.

**TA:**

Cornelius was an astronomer who thought about nothing but astronomy. He was obsessed with stars and galaxies. Gradually he was turning his home into an observatory.

But before he was finished his wife Agatha complained that he was spending too much time renovating the house and not enough on his job which was astronomy. As a result, the house was no longer livable. Cornelius decided that Agatha was right and he went back to thinking about nothing but astronomy.

Then, one evening when Cornelius was walking through his house, an unfinished floor collapsed under his weight. Cornelius was severely injured but the accident showed him that he should try to balance his time between astronomy and maintaining his home.

**MAAO**

There was once a king in a very rich country. He had fifteen armies. Unfortunately, the king was severely injured in a coach crash. His wife was with him at the time. His coach crashed on a bridge.

**SET 14**

**BASE Story:**

Frederick the shepherd was tending his flock at the edge of a forest. Suddenly, a lion appeared out of the forest. Frederick was really scared so he started shouting to scare the lion away. But the shouting did nothing but attract the lion's attention and it began walking towards him. He then realized he should have left it alone.

To escape, Frederick gave the lion a lamb. While the lion was occupied, he led the rest of the flock away. Eventually, he guessed they were far enough away to be safe from the lion. So he took a nap on the grass.

However, when he awoke he noticed another lamb was missing and he could not find it anywhere! At the next opportunity, Frederick bought a blunderbuss so he could better defend his flock in the future.

**FA:**

Zebediah, the old miser, was sitting on his back porch counting his money and he dozed off. When he awoke he discovered that one bag of coins was missing. He could not find the bag anywhere. Zebediah figured that the thief was still nearby so he started yelling for whoever took the money to show his face.

Suddenly, a beggar came out of the forsythia bushes and started walking towards the old man on the porch. Zebediah was surprised and very frightened. He wished he had left the beggar alone.

To escape, he threw a second bag of coins to the beggar. While the beggar was occupied, Zebediah took the last few bags of money into the house and locked the door!

**TA:**

Zebediah, the old miser, was sitting on his back porch counting his money. Suddenly, a beggar came out of the forsythia bushes. Zebediah was very frightened so he started yelling for the beggar to go away. The yelling only caught the beggar's attention and he began walking towards the old man. The miser wished he had left the beggar alone.

To escape, Zebediah threw a bag of coins to the beggar. While the beggar was occupied, Zebediah took the last few bags of money into the house and locked the door.

Zebediah figured that he was safe inside and he dozed off in an easy chair. When he awoke he was surprised to discover another bag of coins was missing. He could not find it anywhere!

**LS:**

Zebediah the cow herder was watching over his herd at the edge of the woods. Suddenly, a monstrous tiger came out of the woods. Zebediah was very frightened so he started yelling for the tiger to go away. The yelling only caught the tiger's attention and it began walking towards the cow her. He wished he had left the tiger alone.

To escape, Zebediah gave the tiger a calf. While the tiger was occupied, Zebediah took the remainder of the herd to a safe place. He figured he was safe and dozed off. When woke he was surprised to discover another calf missing. He could not find it anywhere!

**MA:**

Zebediah the cow herder was watching over his herd and he dozed off. When he awoke he discovered that one calf was missing. He would not find it anywhere. Zebediah figured that the thief was still nearby so he started yelling for whoever stole the calf to show his face.

Suddenly a monstrous tiger came out of the woods and started walking towards him. Zebediah was surprised and very frightened. He wished he had left the tiger alone.

To escape, he gave the tiger a second calf. While the tiger was occupied, Zebediah took the remainder of the herd to a safer place!

**MAAO**

A shepherd walked his sheep through the forest. His flock was quite small. He had twelve lambs. He had heard rumours of lions in the forest. He always took a nap on the grass at the other side of the forest.

## SET 15

**BASE Story:**

Mark the chicken farmer was standing in the henhouse collecting eggs while his chickens were out in the yard. When his chickens returned, they all began to attack him, for some unknown reason.

"You ungrateful beasts!" Mark screamed. The chickens chased him out of the henhouse and into the woods.

While they were gone, a thief slipped into the yard and made off with the unguarded eggs. After he got out of the hospital, Mark quit chicken farming and went on a vacation in Africa.

**FA:**

A tidal wave hit a girl scout camp on the beach while everyone was away and washed away all of the sleeping bags. A short time later, Linda, the camp director, came back and discovered the tragedy.

She stood on the beach thinking about what to do. While she was thinking, the girl scouts returned from hiking in the hills. Since they could not find their sleeping bags they attacked their director. What'll we do now? Where will be sleep?...they said.

"Why do you pester me when I am racking my brain to figure out what we are going to do?" the camp director screamed as she ran down the road. "Show a little appreciation!"

**TA:**

Linda, the director of a girl scout camp, was thinking about the camp on the beach while all the girls were hiking in the hills nearby. While she was standing there the girl scouts returned and they all started pestering her with questions about what was for dinner, could they go swimming, and so on.

"Why do you pester me when I am racking my brain to figure out what we're going to do?" Linda screamed as she ran off the beach and down the road. "Show a little appreciation!" The scouts followed her, still yelling.

A short time later, a tidal wave hit the beach and washed away all the sleeping bags.

**LS:**

Lyle the poultry raiser was thinking about his nests while his poultry were out in the field. While he was thinking the poultry returned and they all started to peck him, for some unknown reason.

"You thankless monsters!" Lyle screamed. The poultry chased him down the road and into the forest.

A burglar came into the garden while they were away and stole all of the nests.

**MA:**

A burglar came into the barn of Lyle the poultry raiser while Lyle was away and stole all the nests. A short time later, Lyle came home and discovered what had happened.

He stood in the barn thinking about what to do. While he was thinking, the poultry returned from the field. Since they could not find their nests they began to peck Lyle.

"Why do you attack me when I am racking my brains to think how I am going to get your nests back?" Lyle screamed as he ran into the forest. "You thankless monsters!"

**MAAO**

A chicken farner was proud of his free range eggs. He was critical of the factory farm in the town. His own eggs sold very well. The factory farm used computers to try and produce more eggs. The farmer liked to go to Africa on vacation.

**SET 16**

**Story:**

Two pioneers, Smith and Johnson, discovered a gorgeous little valley that was as yet unclaimed in Oregon Territory. Each of them wanted the land for himself, so soon they were fighting over who saw it first.

While they were fighting over it, Sheriff Jones came along and divided the land evenly between them. In the end, Smith and Johnson realized that they should have done that on their own. From that day forward they were the best of friends, and they helped build each other's houses.

**FA:**

Betty and Norma were walking towards each other when they simultaneously discovered a twenty dollar bill on the floor. Both of them wanted the money so soon they were arguing desperately over who saw it first.

While they were busy arguing over the money, Lee came along and took it herself. By the time Betty and Norma realized they should just divide it up, it was too late.

**TA:**

Betty and Norma were walking towards each other, when they simultaneously discovered a twenty dollar bill on the floor. Both of them wanted the money so soon they were arguing desperately over who saw it first.

While they were busy arguing, their teacher, Mrs Lee, came along, took the money and divided it between them. Then Betty and Norma realized that they should have done that themselves.

**LS:**

Baggle and Norton were settlers who discovered a beautiful canyon out west, which no one had settled yet. Both of them wanted the canyon, so soon they were arguing desperately over who saw it first.

While they were busy arguing over the canyon, Constable Leigh came along and divided it between them. Then Baggle and Norton realized they should have done that themselves.

**MA:**

Baggle and Norton were settlers who discovered a beautiful canyon out west, which no one had settled yet. Both of them wanted the canyon, so soon they were arguing desperately over who saw it first.

While they were busy arguing over the canyon Constable Leigh came along and took it himself. By the time Baggle and Norton realized they should just divide it up, it was too late.

**MAAO**

Two explorers discovered a new valley in Oregon. There were a lot of trees in the valley. They went to the sherrif to register their claims. They cut down some of the trees whilst building their houses.

**SET 17**

**BASE Story:**

Once there was a teacher named Mrs Jackson who wanted a salary increase. One day, the principal said that he was increasing his own salary by 20 percent. However, he said there was not enough money to give the teachers a salary increase.

When Mrs Jackson heard this she became so angry that she decided to take revenge. The next day, Mrs Jackson used gasoline to set fire to the principal's office. Then she went to a bar and got drunk.

**FA:**

McGhee was a sailor who wanted a few days of vacation on land. One day McGhee became so impatient that he tried to blow up the captain's cabin using dynamite.

After this incident, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on board to repair the ship.

**TA:**

McGhee was a sailor who wanted a few days of vacation on land. One day, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on the ship.

After McGhee heard this he became so upset that he decided to get revenge. Within an hour McGhee blew up the captain's cabin with dynamite.

**LS:**

Professor Rosie McGhee very much wanted a raise. One day the provost announced that he was giving himself a raise. However, he said that since money was short, no one else would get a raise this year.

After Professor McGhee heard this she became so upset that she decided to get even. One hour later, Professor McGhee blew up the administration building with dynamite.

**MA:**

Professor McGhee very much wanted a raise. One day she became so impatient that she used kerosene to burn down the administration building.

After the fire, the provost announced that he was giving himself a raise. However, he said that due to the fire, there was not enough money to give one to anyone else.

**MAAO**

A teacher once thought that she deserved a pay rise. She asked the principal when her rise was due. She was wearing her best suit. The principal told her that rises were decided by the governers.

## SET 18

**BASE Story:**

A cobra called Pierre was slithering through the brush when a falcon came towards it. Pierre became frightened and struck, wounding the falcon in the abdomen. To teach him a lesson, the falcon grabbed the cobra and flew away with it.
Unfortunately for both of them, after they were high above the earth the wounded falcon became too weak to fly. Consequently, they plummeted straight down to the earth and perished. They landed, strangely enough, in the middle of an outdoor wedding ceremony.

**FA:**

Young Gene was walking home from school when he met a policeman. Suddenly, the policeman grabbed him and drove away with him.

Once they got onto the freeway, Gene panicked and hit the policeman in the side of the head. Unfortunately, the policeman became too dizzy to drive. Consequently, they ran off the road and died when they hit a wall.

**TA:**

Young Gene was walking home from school when a policeman approached him. Suddenly Gene panicked and hit the policeman in the side of the head. The policeman wanted to set the boy straight. He grabbed Gene and drove away with him.

Unfortunately, once they were on the road the policeman became too dizzy to drive. Consequently, they ran off the road and died when they hit a wall.

**LS:**

Gene the rattlesnake was going across a field when it met a raven. Gene panicked and bit the raven in the stomach. The raven wanted to set the rattlesnake straight, so it grabbed the snake and took to the air with it.

Unfortunately, once they were in the air, the bleeding raven became too faint to continue. Consequently, they both fell from the sky and died when they hit the rocky ground below.

**MA:**

Gene the rattlesnake was going across a field when it met a raven. Suddenly, the raven grabbed the snake and took to the air with it.

Once they were in the air, Gene panicked and bit the raven in the stomach. Unfortunately, the bleeding raven became too faint to continue. Consequently, they both fell from the sky and died when they hit the rocky ground below.

**MAAO**

Pierre the cobra once met a falcon in a forest. The falcon was called Tarquin. Pierre asked Tarquin to help him get to a wedding on time. Tarquin said "Why not?". He was wearing his flying goggles at the time.

**SET 19**

**BASE Story:**

A twenty year old lady named Samantha lived in a southern state where there was a freak blizzard during the summer with subzero temperatures. For a week, she could not leave her home since she did not have a warm jacket.

Then a meteorologist informed Samantha that the cold days were over. She was very pleased to hear that. However, Samantha was well aware that meteorologists often make mistakes. The first thing she did was sell her summer clothes and buy a warm jacket. She never had a chance to use the jacket again, so she let her kitten sleep in it.

**TA:**

Harold lived in a country frought by civil war. All around his village there were battles. Because he did not have a gun he could not leave the village for several days.

Before long, a soldier told Harold that the war was over. He was delighted to hear that. But he was well aware that the soldier might be wrong. Immediately he sold his season ticket to the symphony and bought a handgun.

**FA:**

A soldier told Harold that the war was over. Harold was delighted to hear that. It never occurred to him that the soldier might be wrong. Immediately, he sold his handgun and bought a season ticket to the symphony.

But things worked out badly. Before long, battles sprang up all around the village. Because he had no gun, Harold could not leave the village for several days.

**LS:**

Carol lived in a southern country where there was a terrible, freezing snowstorm. Because she had no coat, Carol could not leave her apartment for several days.

Before long, a weather forecaster told Carol that the winter weather was over until next year. The young woman was delighted with the news. But she was well aware that the weather forecaster could be wrong. Immediately she sold her lighter clothes and bought a fur coat.

**MA:**

A weather forecaster told Carol that the winter weather was over until next year. The young woman was delighted with the news. It never occurred to her that the weather forecaster could be wrong. Immediately, she sold her fur coat and bought lighter clothing.

But things worked out badly. Before long, there was a terrible, freezing snowstorm. Because she had no coat Carol could not leave her apartment for several days.

**MAAO**

Once a freak blizzard hit the southern states. There was snow on the ground in Kentucky in August. People wore their winter clothes for a week. The radio told people when the weather was about to hit.

**SET 20**

**Story:**

Chester Inc was being pursued by the Securities Exchange Commission for keeping improper records. Therefore, Chester enlisted the help of Honest Abe Accountants to protect its interests. Honest Abe showed Chester Inc how to conceal its errors. But Chester Inc recognized it should not rely on Honest Abe to cover for it. Hence, Chester fled from the town without paying the accounting firm.

Then the auditor arrived. Under questioning, Honest Abe issued a statement that was vague but not exactly dishonest in an attempt to protect Chester Inc. Luckily for Chester, the auditor was convinced and a short time later, he cancelled the audit.

Still, Honest Abe Accountants were insulted that Chester would not pay, and the accountants brought a suit against the corporation.

**TA:**

A bank robber known as Arnie was running from a constable when he met a gardener working in a yard. Arnie begged the gardener for help. The gardener told him to hide behind a tree, which he did. However, Arnie considered the situation and realized he should not trust the gardener. With this in mind, he slipped away into the forest without thanking the gardener.

In no time the constable arrived and asked the gardener if he had seen a robber. The gardener answered that he was busy gardening which was not really a lie. As it happened, the constable did not notice anything odd and soon gave up the search.

**FA:**

A bank robber known as Arnie was running from a constable when he met a gardener working in a yard. He begged the gardener for a place to hide. The gardener told him to hide in the garage, which he did.

In no time a constable arrived and asked the gardener if he had seen a robber. The gardener answered that he hadn't seen anyone. However, at the same time he pointed at the garage. As it happened, the constable did not notice anything odd and soon gave up the search.

Arnie considered what had happened and realized that he could not trust the gardener. With this in mind, he slipped away into the woods without thanking the gardener.

**LS:**

Acme Co. had cheated on its taxes and was running from the IRS when it hired a financial consultant for help. The financial consultant told Acme how to hide its mistakes. But Acme realized it should not trust the consultant. With this in mind, Acme slipped out of the city without compensating the consultant.

In no time the IRS agent arrived and asked about Acme's taxes. The consultant answered with a memo that covered things up but was not really a lie. Fortunately for Acme Co., the IRS agent did not notice anything odd and soon dropped the investigation.

**MA:**

Acme Co. had cheated on its taxes and was running from the IRS when it hired a financial consultant for help. The financial consultant told Acme how to hide its mistakes.

In no time the IRS agent arrived and asked about Acme's taxes. The consultant answered with a memo that sounded smooth. However, it purposely hinted at Acme's transgressions. "My client is not a crook... besides, one can find discrepancies in the books of any company." Fortunately, the IRS agent did not notice anything odd and soon dropped the investigation.

Acme Co. considered what had happened and realized it could not trust the consultant. With this in mind, Acme slipped out of the city without compensating the consultant.

**MAAO**

A company once got into trouble for keeping poor accounts. There managing director was called Chester. He wasn't very good at book keeping. After the accountants had sorted things out, he went on holiday for a few weeks.

## *Appendix B: Materials Used In Experiment 3*

**Set1**

**B**

The Ralua is a tall, blue plant which used to be widespread throughout the country. For some reason, in the last few years many of its seeds have been attracted to the roadside, where they try to grow.

However, the plants are not able to absorb all the poisonous gases from the traffic very well, which has serious consequences.

They produce fewer seeds for reproduction every year and also tend to die fairly early.

If this trend continues, the Ralua will be extinct within the next 200 years.

**LS**

The Havanna is a tall plant, which used to be large in number throughout Hoholand. It is not known why, but for the last few decades many of its seeds have been attracted to settle near a herb called kali, which makes them grow bigger.

However, there are some serious side effects.

They become less reproductive and die at an earlier age. If this trend continues, the Havanna plant will not be around in the next century.

**SSO**

The Havanna are a tribe living in central Gunaland. They are an isolated tribe which used to be very large in number. Several decades ago they discovered a herb called gabon which, when eaten, made them grow bigger.

They became addicted to it but there were serious side effects.

Their sex-drive decreased, leading to fewer couples having children. As well as this, it was common for more heart attacks to happen and at an earlier age. Their number is therefore decreasing to such an extent that if the trend continues, the Havanna will not be around to see in the next century.

**MA**

The Havanna is a tall, blue plant, which used to be large in number throughout Hoholand. In the past few years, they have commonly been found near the roadside, where their leaves become a darker shade of blue, due to the chemicals they find there.

If they keep growing by the roadside, the Havanna will become a navy blue in appearance within the next 200 years.

**SMAO**

The Havanna are a tall, dark-haired tribe who were once commonly found in central Gunaland. In the last few years, many Havanna have moved to be near the herb gabon, which makes them grow taller when eaten.

If they keep eating kali, the Havanna will be the tallest tribe around in the near future.

**OO**

The Ralua is a well-known tall, blue plant which used to be common throughout the country.

Many of its seeds are green. A Ralua makes a lovely indoor plant.

**Set 2**

**B**

The Siri live in a country governed by Al-Siri. Al-Siri and his family enjoy all the benefits of their power and position in society, with everything being done for him.

It is tradition that the strongest son of this blood-line inherits the throne.

Al-Siri, being the eldest son, does not need to do any work as his people provide him with a constant supply of the finest food and clothes. The people of the country keep doing this as they know they will face terrible consequences if they don't keep him happy, as they will be burned alive.

**LS**

King Balu is leader of the Bali people, who live in a warm country known as Futu. The king and his immediate family are provided with all the necessities they require, and are looked after well by the other Balis.

It is custom that to become king, you have to be the fattest son of the previous one.

King Balu, being the largest son, is provided with a constant supply of riris and holos by his people to keep him happy, and does not need to do any work. His countrymen do this as they have been warned that if they don't find enough riris and holos, their cannibalistic king will eat them.

**SSO**

The Kula species, who live in the warm waters of Futu, in that each shoal has one large male member called a Kembali. The Kembali is protected and looked after by the remaining Kula.

The Kembali is the largest of the male Kula species, and is the only male to reproduce.

The Kembali does nothing, except eating small riris and holos which the Kula swim around and search for, in order to keep him well fed. The Kula do this as they know that if they do not find enough food, the Kembali will eat them.

**MA**

The Bali live in a country governed by Al-Balu. Al-Balu and his family's every need is catered for and they enjoy their position in society.

As is tradition, the fattest Al-Balu gets to sit on a golden chair by being the leader.

The only obligation on the Al-Balu is to lead his people in invasions of neighbouring countries when land becomes scarce.

**SMAO**

The Kula species have a kind of leader-figure known as a Kembali. The species is found in the warm waters of the ocean. The Kembali is well looked after by the other Kula.

Kembalis get to live in a special cave in a coral reef.

When food in their waters becomes scarce, the Kembali leads the rest of the shoal in attack on the waters of neighbouring colonies of fish, to take over their territory.

**OO**

A certain king is the leader of the Siri people, who live in a warm country.

His people like lots of food and clothes.

Sometimes he will burn the food and clothes and buy new ones.

**Set 3**

B

Snordon is played on a square board divided into sections. Two lagnogs, or players, take turns to move boli (small disks) around the board.

When an individual bolum becomes surrounded by the other player's boli, it changes ownership and joins the surrounding group.

When all of the boli are in possession of one lagnog, the game is over.

**LS**

Blagson is played on a rectangular board divided into vodogs. Two galps, or contestants, take turns to move 11 small cubes around the board.

Should a single cube become surrounded by the other galp's cubes, it changes ownership and joins the surrounding group.

When all the cubes become the possession of one galp, blagson is over.

**SSO**

Blagson is carried out on a large flat field divided into vodogs. Two balgs, or teams, of 11 horsemen are required.

The horsemen manoeuvre for position throughout blagson.

When one horseman becomes isolated and encircled by those of the other balg, he joins the opposition and becomes part of the encircling balg.

When one balg has gained possession of all the horsemen, blagson is over.

**MA**

Blagson is played on a board divided into vodogs. During blagson, galps, or contestants, move small cubes around the board.

If a cube moves onto the same position as that of one of the opponent's cubes, both players have a drink of vodka.

The game continues until one player gets bored or too inebriated to continue.

**SMAO**

Blagson is carried out on a field that is divided into vodogs. Two balgs, or teams, take it in turn to move their horsemen around the field.

If one horseman moves into the identical part of the field as an opponent, both players wave their flags.

Blagson continues until one team gets bored, or their arms become too tired.

**OO**

Snordon is a well-known game. It is played in pubs and clubs.

The pieces are made from halban. Some snordon boards are very valuable.

**Set 4**

**B**

Monominos live around Lake Wanubi on the fringes of the Western Sahara. A monomino has hard skin and is about the size of a small pig. It sleeps during the day and comes out at night, when it searches for food.

Male monominos will not allow other males into their area. They warn rival males off by rubbing their backs against trees, which gives off a warning scent. Should its area be invaded by another monomino, they will fight until there is only one survivor.

**LS**

Rambos live near Lake Zuvius on the outskirts of the Western desert. A rambo has tough claws and is the size of a small elephant. They rest when it rains and hunt when it is dry.

Male rambos do not permit other males into their territory. They try to scare other males away by scratching a warning mark on trees with their claws. If another rambo should enter the territory, they will start fighting, with only one rambo staying alive.

**SSO**

The mountains of Zuvius are inhabited by the Chang tribe. They carry sharp knives are about as tall as a fully grown elephant. They sleep when it rains and only hunt for food when it is dry.

The Changs do not want any other tribe eating their crops. They try to scare other tribes off by carving crosses high on trees, which is seen as a warning symbol. If a different tribe tries to eat some of their crops, they have to fight against the Changs until one tribe is wiped out.

**MA**

Rambos live near Lake Zuvius, situated on the fringes of the Western Sahara. They have tough skin and are the size roughly of a small elephant.

They tend to avoid rain at all costs. When they get lonely, rambos call out with a plaintive cry. When another rambo enters its area, they will tend to play together.

**SMAO**

The Chang tribe lives in the vicinity of Lake Zuvius, just outside the desert. They are tall, about the size of a small elephant and carry sharp knives. They tend to sleep when it rains, but don't like sharing their food with other Changs.

Changs are not fond of bathing. When they get bored, the Changs sing a plaintive tribal song. When another tribe enters their area, they will play a primitive form of football together.

**OO**

A mysterious group of pig-sized animals live near the desert. They sleep during the day and hunt at night time. They have extremely hairy legs.

**set 5**

**B**

Karla is a novel type of cooking pot, used by the Timuni in Alnata.

The structure of the Karla is designed in order to reduce the heat inside, and therefore prevents the food getting burned in the scorching cooking fires.

Water is poured into a layer of the Karla during cooking, which cools the food.

**LS**

The Valkri is a special kind of frying pan, used by the Jalpeni in Frodon.

The Valkri is created in such a way as to be able to reduce heat, thereby preventing meat being getting burned when using the extreme temperatures of the cooking fires.

A liquid is poured into the layers of the frying pan when cooking, which cools the temperature of the meat.

**SSO**

The Vubu is a special wall built by the Jakar tribesmen in Frodon.

The Vubu is built in such a way as to be able to reduce the heat within it, thereby preventing the Jakar from sweating too much in the extreme temperatures of the midday sun.

A liquid is pumped through the Vubu, which cools the stone and therefore prevents the Jakar within the walls from getting too hot.

**MA**

The people of Frodon use a special type of frying pan, known as the Valkri.

The structure of the Valkri is specially designed in order to allow it to be handled by children, as this can be difficult.

It has a U-shaped handle, which enables it to be held by people with small hands.

**SMAO**

The Jakar tribesmen of Frodon have built a special wall known as the Vubu.

The Vubu's stone gates can be opened by elderly people, despite their heavy weight.

Handles set in the wall incorporate springs, which allow weaker people to open the gates.

**OO**

A new type of cooking pot, called the Karla, is used by the people of Alnata. Karla's can be purchased in a range of colours. Food cooked in a Karla tastes great.

**Set 6:**

**BASE:**

The podash is an animal prized by the Zad of Aluck. Podash are coveted because of their wonderful property of being able to travel large distances, carrying very large loads, whilst eating relatively little. Obviously this makes them extremely cheap to keep, despite the huge numbers of chickens, or sacks of rice that the Zad barter them for. Most podash live to an average age of 8 years, during which time they serve their owners extremely well.

Unfortunately, podash are not terribly common, despite their desirability. It takes a female 18 months to conceive, and even then, the child might be deformed. Hence the scarcity of these creatures.

**LS:**

A terin is a type of animal held in great esteem by the Norin tribe of Itran. They are desirable because of their incredible stamina whilst transporting goods across vast distances, on very little food, which makes them very economical, regardless of the fact that they are traded for vast sums. The life expectancy of the terin is around 8 years, throughout which they are a real investment for their owners.

Annoyingly, terin are not very easy to come by, despite their value. This is because female terin are pregnant for 18 months, and there is no guarantee that the offspring will be healthy. Therefore they are quite rare.

**SSO:**

Sagi are the revered priests of the Norin tribe, which can be found in NE Itran. These people are have fantastic stamina and can travel for days without water or sustenance, in the baking heat, earning them the respect of the Norin people. This obviously makes them extremely useful as messengers to other tribes, despite the huge amounts of food and water they require for their services. Most Sagi have an extremely long life expectancy, which means they can cross the deserts, for the tribe, for many years.

Sagi are very rare, however, despite the fact that many people aspire to their calling. This is because it takes a long time to train to become a full blown priest, and even then they might not pass the requirements. For these reasons, the Sagi are not often found.

**SMAO:**

Sagi are the revered priests of the Norin tribe, which can be found in NE Itran. This country is very hot, and many people cannot go there, on account of the baking heat. The Sagi have fantastic stamina and can travel for days without water or sustenance. Since the Sagi are devout priests, they worship all the time.

Sagi are very rare, however, despite the fact that many people aspire to their calling. Many people have tried to document the lifestyle of the Sagi, but this has proven too difficult for most. They are so rare because to become a priest you have to bathe everyday in cold water. Most people find that this puts them off after a week or two.

**MA:**

A terin is a type of animal held in great esteem by the Norin tribe of Itran. This is a relatively uninhabitable land, but the Norin guard their territory jealously. Terin have an amazing sense of smell, and can find scent trails made even weeks beforehand. Since terin are clean animals, they wash themselves all the time.

Many people have tried to film the terin but few have succeeded. They are quite uncommon because in order to mate the terin must fight violently first. This puts most of the animals off mating after one or two attempts.

**OO:**

There is an incredible animal called a podash. It is used to carry messages. They can be hard to find. Podashes are rather smelly, but they do like children

**Set 7:**

**BASE:**

Frent is a chant used by the Syman of Lowari, during tribal feasts. The chant is said in pairs, either single or mixed sex, and before the couple can start the families of both parties must meet to agree on the people who will say Frent. The sex and marital status of the possible candidates determines who can chant Frent. This is because the chant is believed to promote inter-family relationships and therefore it would be indiscreet for two married people to chant together.

Frent is only chanted by the two people chosen, immediately before the tribal feast. This seems strange at first, but this tradition developed due to Frent's very complicated nature: it would be impossible for a couple to remember it after their heads become fuddled by a few drinks.

**LS:**

Boga is a chant used by the Vari of Musuka, whilst having tribal meals. The chant is recited by a single couple, in which the participants can be of one or both genders, and before the pair can begin the relatives of both must gather in order to establish who should perform Boga. The gender and matrimonial history of the possible chanters helps show who may say Boga. The reason behind this is that the chant is held to lead to relationships between the families and therefore it would be improper for a couple to speak with each other, if they were both married.

Only the ordained couple can recite Boga, which they do just before the ritual meal. This seems a little strange until you remember how complicated Boga is: couples who to have a drink beforehand find that their dazed wits fail to remember the chant properly.

**SSO:**

Children in Vari-Musuka schools frequently play the game of agabu before noon. Both sexes are allowed to play, either with members of the same sex, or members of the opposite sex. Only one pair plays at a time. Before the game can be played, the members of each gang from which the opponents come must agree who will play agabu. This depends on the size of the potential players, and their level of expertise. The reason behind this is that the players must be evenly matched, since this game determines who will play together in the playground.

The opponents play on their own, in a shady area of the playground. This seems a little strange until you remember how complicated agabu is: opponents who sit out in the midday sun to play become so dazed by the heat that they cannot remember the rules properly.

**SMAO:**

Children in Vari-Musuka schools frequently play the game of agabu. This game was established many years ago, by rival gangs, who wanted to decide who got to be first in the lunch queue. If you lost, you got to go first. No-one sex really wants to play, but each gang forces a couple of people to take part out of tradition.

Agabu is played in a shady area of the playground. The last person playing agabu has lost. As it is hard to establish who actually was last, the gangs tickle the players until they tell the truth.

**MA:**

Boga is a chant used by the Vari of Musuka, whilst having tribal meals. It originated with the original tribe of the area, who couldn't be bothered with fighting, and found that singing together was much more fun and productive. No family in the tribe is particularly keen to chant, but each family nominates a few participants.

Boga is said sitting at the table. The person who finishes chanting Boga last is loses. As it is hard to tell who this was, the families tell the chanters really bad jokes until the loser confesses.

**OO:**

There is a type of chant that certain tribes. Two people say it before a meal. It is said at a table. Meals in this tribe consist of maize, beef and biri-biri wine. The chant can sound quite boring.

**Set 8**

**BASE:**

Ablua is a type of food eaten by the Amorki of Nashak. Ablua is made from boiling the pulp of Ana fruits, which are crushed and drained first. It has a quite distinct flavour, being quite light and scented. It is quite soggy and mushy, and it is a sunshine colour. It is traditionally eaten at weddings and funerals, being a great delicacy.

It is said that Ablua can enhance your attractiveness, and many people smear it on their skin. Science has not yet proven this to be true, but it certainly acts as a natural perfume.

**LS:**

Voya is a type of food that the Zubu of Itlad eat. To prepare Voya, the flesh of Chup fruits is simmered, having been pulped and left to drain previously. Voya has a very unique taste, being quite delicate and fragrant. It is reasonably sticky and soft, and is coloured yellow. Since it is considered such a treat, it is usually only consumed at special occasions. It is said that Voya can increase your powers of attraction, and many people rub it into their skin. Tests haven't yet showed the truth of this, but it definitely makes people smell better.

**SSO:**

The drink Voya is drunk mainly by businessmen in Imron. The drink itself is made by fermenting and distilling the husks of Chup seeds, which are thoroughly washed and dried. Voya is quite a pleasant drink, and is quite acidic to taste. It is quite sticky and fibrous, containing sediment at the bottom. Being so expensive, it is usually drunk at business dinners, where people want to impress their clients.

It is said that Voya can aid digestion, if drunk regularly. Tests haven't yet showed the truth of this, but it certainly adds to the enjoyment of one's meal, because it makes food taste better.

**SMAO:**

The drink Voya is drunk mainly by businessmen in Imron. These businessmen are generally quite pompous and do not realise how annoying they are. They find that Voya relaxes them whilst doing their accounts. It is considered a great honour to do accounts in Imron.

It is said that Voya can aid digestion, but no-one knows if this is true. It contains quite a lot of sediment and can occasionally be quite opaque. It is, moreover, quite sticky, and tends to make your breath smell, which is why the businessmen's wives leave them alone.

**MA:**

Voya is a type of food that the Zubu of Itlad eat. Zubu are quite interesting people, and are quite unaware of the spectacle caused by their small stature. Voya is eaten to increase their energy whilst the Zubu tend the fields, a job held in great esteem.

It is said that Voya is a good cure for hangovers, but no-one is sure this is true. Its smell is vaguely reminiscent of clover, and can occasionally smell strongly like honey. Unfortunately it stains clothes and tends to cause halitosis, which is why other tribes avoid the Zubu.

**OO:**

Ablua is a type of food. It has quite a unique smell, and is reasonably attractive. It makes a wonderful Christmas present. It is easy to overindulge when eating Ablua.

................

# *Appendix C: Materials Used In Experiment 4*

Bases, targets and inferences used in Experiment 4. The A set of stimuli (where A type inferences receive structural support) is followed by the B set, where structure supports B type inferences (in the experiment presentation order was randomised).

**Chateau Bogusse:**
is a vineyard.
is in the southern french district of Pretence.
has sandy soils, with a lot of surface pebbles
has a warm microclimate which enables grapes that are described in French as *elanger* to be produced.
**Domaine Fraudulent:**
grows plums.
has clay soils in which wildflowers grow
is in the western Departement of Maidoop.
its fine microclimate causes plums that are described in French as *moublier* to grow.
Domaine Fraudulent's *moublier* plums are prized and sell for high prices.
**Mas de la Fiction:**
grows melons.
is in the southern Departement of Whaupper.
has sandy soils, with a lot of surface pebbles
its warm microclimate causes melons that are described in French as *jayert* to grow
Mas de la Fiction's *jayert* melons are held in low esteem, and sell poorly.

A Chateau Bogusse's grapes are highly prized and sell for high prices
B Chateau Bogusse's grapes are held in low esteem, and sell poorly.

**Helpful Dictionary**
**Elanger:**
is a slang (argot) term used by french farmers
the term is thought to have originated in Belgium
is a way of describing extremely ripe fruit.
the extreme ripeness causes some of the moisture in the fruit to evaporate.
this evaporation concentrates the fruits' flavours
**Jayert:**
is a slang (argot) term used by french farmers
the term is thought to have originated in Belgium
is a way of describing extremely ripe fruit.
the extreme ripeness causes the fruit to become very sweet
this super-sweetness makes the walls of the fruit soft and squashy
**Moublier:**
is a technical term used in french agricultural colleges
the term is thought to have originated in Spain
is a way of describing extremely ripe fruit.
the extreme ripeness causes some of the moisture in the fruit to evaporate
this evaporation leads to extremely concentrated flavours

**SCENARIOS**
**BASE - The Guralaga**
can be found in Australia
lives in Rainforests
only eats gau-gau berries
has a cronomus lucundus
the cronomus lucundus enables the Guralaga to eat gau gau berries.
**TARGET 1 - The Mongret**
can be found in Australia
lives in Rainforests
only eats gau-gau berries
has a probus razoris
the probus razoris enables the Mongret to eat the gau gau berries.
Thanks to the way they eat, Mongrets live to a ripe old age and rarely suffer from cancer
**TARGET 2 - The Crany Dog**
can be found in Papua new Guinea
lives in the grassy backlands
eats vegetation
has a remulum grandoso
because of the remulum grandoso the Crany Dog can eat vegetation.
Crany Dogs are particularly prone to cancer, which originates in their digestive system.

A Guralaga are particularly prone to cancer

B Guralaga live to a ripe old age and rarely suffer from cancer

**"DICTIONARY ENTRIES"**
**BASE DICTIONARY ENTRY**
Cronomus lucundus
are unique to certain types of bird
are important to berry eaters
is a long spleen-like organ
keeping berries in the cronomus lucundus allows the berries to slowly ferment, allowing the goodness inside the bitter skins to be released
**SMT DICTIONARY ENTRY**
Probus razoris
are unique to certain types of bird
are important to berry eaters
is a long plier-like bill
crushing berries in the probus razoris allows the goodness inside their bitter skins to be released without the skins having to be swallowed
**SST DICTIONARY ENTRY**
Remulum grandoso
is unique to certain types of dog
are important to dogs which eat a wide range of vegetation
is a short intestine-like organ
keeping vegetation in the remulum grandoso allows it to slowly ferment, allowing the goodness inside the outer skins to be released

**SCENARIOS**
**BASE - Lexicon House**
Is over two hundred years old.
Is owned by a hotel group.
Has floor to ceiling windows.
the construction techniques used in Lexicon House prompted a leading architecture expert to describe it as a Scotland's finest example of "flamberge"
**TARGET 1 - Grange Manor**
Is over two hundred years old
Is owned by a hotel group.
Has floor to ceiling windows.
The way the Manor is built leads architectural experts to describe it as a good example of the "greaves" style.
Grange Manor is slowly crumbling due to the way it is built, and requires a great deal of maintenance
**TARGET 2 - The Old Manse**
Has been painted in lilac recently.
Is owned by the Earl of Murray.
has delightful pillars.
The Manse's construction made architectural experts claim the building was a tremendous example of the style known as "crypto -baronic"
The old Manse has needed very few repairs due to wear in its time, and can still be affordably run by its owners

A Lexicon House can still be affordably run by its owners

B Lexicon House requires a great deal of maintenance

**"DICTIONARY ENTRIES"**
**BASE DICTIONARY ENTRY**
**FLAMBERGE**
originates in the Tudor period
is an architectural term
refers to the way walls are constructed
Flamberge buildings traditionally have a lot of duplication in their construction details, a result of defensive measures that were incorporated in their design
**SMT DICTIONARY ENTRY**
**GREAVES**
originates in the Tudor period
is an architectural term
refers to the way walls are constructed
Greaves encompasses features incorporated into the design that reflects religious sentiments, since form rather than function was upper most in the minds of the designers of these buildings.
**SST DICTIONARY ENTRY**
**CRYPTO-BARONIC**
originates in the Elizabethan period
is an architectural term.
refers to a distinctive method of supporting the building fabric
Crypto-baronic buildings traditionally have a lot of duplication in their construction details, a result of defensive measures that were incorporated in their design

**Krav Maga**
this martial art originated in China
it is taught in groups
students fight with short sticks
it is hard to learn
When someone has managed to learn krav Maga they can wear a 'ver-krap. '
**Sampa**
this sport originated in Japan
students fight with their hands and feet
Sampa is taught in one to one lessons by a grand master
it takes a long time to learn
When someone has taken the time to learn Sampa they can wear a 'koo-chai. '
Despite it's defensive nature, and perhaps because of the 'koo-chai', students of this martial art get in lots of fights, and are frequently injured.
**Escrima**
this martial art originated in China
students fight with short sticks
it is taught in groups
it is hard to learn
When someone has managed to learn Escrima they can wear a 'rok-ard.'
Despite it's physical nature, and perhaps because of the 'rok-ard', Sampa injuries are rare.

A krav Maga students manage to stay relatively injury free.

B krav Maga students are frequently injured.

**koo-chai**
is made of cotton
comes in a range of colours
is heavily padded
ties with a cord around the wiast
the heavy padding results in a certain loss of mobility for the wearer
**rok-ard**
comes in a range of colours
is made of padded cotton
is decorated with offensive and aggressive symbols
ties with a cord around the waste
the decorative symbols signify that the wearer is entitled to wear a the rok-ard
**Ver-Krap**
comes only in white
is made of wool
the wool is wrapped around the wearer in many layers
has a nifty drawstring to secure it in place
the many layers of wool result in a certain loss of mobility for the wearer

**The Floating Bridge**
has a wooden bar
has lots of live music
serves beer brewed on the premises
the beer is made by a secret
method called 'sepim'
**The Albert**
is oak panelled
has loads of live gigs
serves cider brewed on the premises
the cider is made using a
secret process called 'tannon'.
The tannon process makes a horrible smell that makes the Albert's beer garden unbearable because of the strong smells that waft across it
**The Castle Hotel**
has a modern design
has a juke box
serves a locally distilled spirit
the spirit is made by using a
secret technique only known as 'hoffam'
The hoffam technique produces a lot of heat that makes the Castle Hotel's beer garden bearable even in cold weather thanks to gusts of warm air from the distillery.

A The Floating Bridge's beer garden is bearable even in cold weather thanks to gusts of warm air of from the brewery.

B The Floating Bridge's beer garden is unbearable because of the strong smells that waft across it

**Helpful Dictionary**
**Sepim**
is a process that aids fermentation
it was perfected in Holland
a secret ingredient is added to the must
the ingredient causes the must to react vigorously, increasing the rate of reaction therein
**Tannon**
is a process that aids fermentation
it was perfected in Holland
a secret ingredient is added to the must
the ingredient causes the yeasts to give off a great deal gas, which adds a distinctive flavour to the brew
**Hoffam**
is a process that aids distillation
it was perfected in Scotland
the mash is kept in a special copper drum.
the copper causes the mash to react vigorously, increasing the rate of reaction therein

## B type sets:

**Chateau Bogusse:**
is a vineyard.
is in the southern french district of Pretence.
has sandy soils, with a lot of surface pebbles
has a warm microclimate which enables grapes that are described in French as *elanger* to be produced.
**Mas de la Fiction:**
has clay soils in which wildflowers grow
is in the western Departement of Maidoop.
its fine microclimate causes melons that are described in French as *jayert* to grow
Mas de la Fiction's *jayert* melons are held in low esteem, and sell poorly.
**Domaine Fraudulent:**
grows plums.
grows melons.
is in the southern Departement of Whaupper.
has sandy soils, with a lot of surface pebbles
its fine microclimate causes plums that are described in French as *moublier* to grow.
Domaine Fraudulent's *moublier* plums are prized and sell for high prices.

A Chateau Bogusse's grapes are highly prized and sell for high prices
B Chateau Bogusse's grapes are held in low esteem, and sell poorly.

**Helpful Dictionary**
**Elanger:**
is a slang (argot) term used by french farmers
the term is thought to have originated in Belgium
is a way of describing extremely ripe fruit.
the extreme ripeness causes the fruit to become very sweet
this super-sweetness makes the walls of the fruit soft and squashy
**Jayert:**
is a technical term used in french agricultural colleges
the term is thought to have originated in Spain
is a way of describing extremely ripe fruit.
the extreme ripeness causes the fruit to become very sweet
this super-sweetness makes the walls of the fruit weaken and break
**Moublier:**
is a slang (argot) term used by french farmers
the term is thought to have originated in Belgium
is a way of describing extremely ripe fruit.
the extreme ripeness causes some of the moisture in the fruit to evaporate
this evaporation leads to extremely concentrated flavours

**SCENARIOS**
**BASE - The Guralaga**
can be found in Australia
lives in Rainforests
only eats gau-gau berries
has a cronomus lucundus
the cronomus lucundus enables the Guralaga to eat gau gau berries.
**TARGET 1 - The Mongret**
can be found in Australia
lives in Rainforests
only eats gau-gau berries
has a remulum grandoso
because of the remulum grandoso the Mongret can eat gau-gau berries.
Mongret s are particularly prone to cancer, which originates in their digestive system.
**TARGET 2 - The Crany Dog**
can be found in Papua new Guinea
lives in the grassy backlands
eats vegetation
has a probus razoris
the probus razoris enables the Crany Dog to eat vegetation.
Thanks to the way they eat, Crany Dogs live to a ripe old age and rarely suffer from cancer

A Guralaga are particularly prone to cancer

B Guralaga live to a ripe old age and rarely suffer from cancer

**"DICTIONARY ENTRIES"**
**BASE DICTIONARY ENTRY**
Cronomus lucundus
are unique to certain types of bird
are important to berry eaters
is a long plier-like bil
crushing berries in the cronomus lucundus allows the goodness inside their bitter skins to be released without the skins having to be swallowed
**SMT DICTIONARY ENTRY**
Probus razoris
are unique to certain types of bird
are important to berry eaters
is a long spleen-like organ l
keeping berries in the probus razoris allows the berries to slowly ferment, allowing the goodness inside the bitter skins to be released
**SST DICTIONARY ENTRY**
Remulum grandoso
is unique to certain types of dog
are important to dogs which eat a wide range of vegetation
is a long powerful jaw
crushing vegetation in the remulum grandoso allows the goodness inside its nasty skins to be released without the skins having to be swallowed

**SCENARIOS**
**BASE - Lexicon House**
Is over two hundred years old.
Is owned by a hotel group.
Has floor to ceiling windows.
the construction techniques used in Lexicon House prompted a leading architecture expert to describe it as a Scotland's finest example of "flamberge"
**TARGET 1 - Grange Manor**
Is over two hundred years old
Is owned by a hotel group.
Has floor to ceiling windows.
The way the Manor is built leads architectural experts to describe it as a good example of the "greaves" style.
Grange Manor has needed very few repairs due to wear in its time, and can still be affordably run by its owners
**TARGET 2 - The Old Manse**
Has been painted in lilac recently.
Is owned by the Earl of Murray.
has delightful pillars.
The Manse's construction made architectural experts claim the building was a tremendous example of the style known as "crypto -baronic"
The old Manse is slowly crumbling due to the way it is built, and requires a great deal of maintenance

A Lexicon House can still be affordably run by its owners

B Lexicon House requires a great deal of maintenance

**"DICTIONARY ENTRIES"**
**BASE DICTIONARY ENTRY**
**FLAMBERGE**
originates in the Tudor period
is an architectural term
refers to the way walls are constructed
Flamberge buildings incorporate features in their designs that reflect religious sentiments, since form rather than function was upper most in the minds of the designers of these buildings.
**SMT DICTIONARY ENTRY**
**GREAVES**
originates in the Tudor period
is an architectural term
refers to the way walls are constructed
Greaves buildings traditionally have a lot of duplication in their construction details, a result of defensive measures that were incorporated in their design
**SST DICTIONARY ENTRY**
**CRYPTO-BARONIC**
originates in the Elizabethan period
is an architectural term.
refers to a distinctive method of supporting the building fabric
Crypto-baronic encompasses design features that reflect religious sentiments, since form rather than function was upper most in the minds of the designers of these buildings.

**Krav Maga**
this martial art originated in China
it is taught in groups
students fight with short sticks
it is hard to learn
When someone has managed to learn krav Maga they can wear a
'ver-krap. '
**Sampa**
this sport originated in Japan
students fight with their hands and feet
Sampa is taught in one to one lessons by a grand master
it takes a long time to learn
When someone has taken the time to learn Sampa they can wear a
'koo-chai. '
Despite it's defensive nature, and perhaps because of the 'koo-chai',
students of this martial art get in lots of fights, and are frequently
injured.
**Escrima**
this martial art originated in China
students fight with short sticks
it is taught in groups
it is hard to learn
When someone has managed to learn Escrima they can wear a
'rok-ard.'
Despite it's physical nature, and perhaps because of the 'rok-ard',
Sampa injuries are rare.

A krav Maga students manage to stay relatively injury free.

B krav Maga students are frequently injured.

**koo-chai**
comes in a range of colours
is made of padded cotton
is wrapped around the wearer in many layers
is decorated with colourful symbols
ties with a cord around the waste
the many layers of coton result in a certain loss of mobility for the wearer
**rok-ard**
is made of cotton
comes in a range of colours
is made of padded cotton
is decorated with aggressive symbols
is wrapped around the wearer in many layers
ties with a cord around the waste
the decorative symbols signify that the wearer is entitled to wear a
the rok-ard
**Ver-Krap**
comes only in white
is made of wool
is worn over only one shoulder
has a nifty drawstring to secure it in place
is inscribed with offensive pictures
the offensive pictures signify that the wearer is entitled to wear a
the Ver-Krap

**The Floating Bridge**
has a wooden bar
has lots of live music
serves beer brewed on the premises
the beer is made by a secret
method called 'sepim'
**The Albert**
has a modern design
has a juke box
serves a locally distilled spirit
the spirit is made by using a
secret process called 'tannon'.
The tannon process makes a horrible smell that makes the Albert's beer garden unbearable because of the strong smells that waft across it
**The Castle Hotel**
is oak panelled
has loads of live gigs
serves cider brewed on the premises
the cider is made by using a
secret technique only known as 'hoffam'
The hoffam technique produces a lot of heat that makes the Castle Hotel's beer garden bearable even in cold weather thanks to gusts of warm air from the distillery.

A The Floating Bridge's beer garden is bearable even in cold weather thanks to gusts of warm air of from the brewery.

B The Floating Bridge's beer garden is unbearable because of the strong smells that waft across it

**Helpful Dictionary**
**Sepim**
is a process that aids fermentation
it was perfected in Holland
the mash is kept in a special copper kettle.
the ingredient causes the yeasts to give off a great deal gas, which adds a distinctive flavour to the brew
**Tannon**
is a process that aids distillation
it was perfected in Scotland
a secret ingredient is added to the must
the ingredient causes the yeasts to give off a great deal gas, which adds a distinctive flavour to the brew
**Hoffam**
is a process that aids fermentation
it was perfected in Holland
the mash is kept in a special copper drum.
the copper causes the mash to react vigorously, increasing the rate of reaction therein

# Appendix D: Materials Used In Experiment 5

The bases, targets and stimuli used in Experiment 5.

Base and Target Scenarios

Geoff runs a coffee shop in the business district of Nawatobee, a small city in the American mid-west. The coffee shop is called MacCawber's. It is owned by Mr Fingus MacCawber, a fierce fastidious man who owns six other businesses in the greater Nawatobee area. Mr MacCawber visits each of his businesses for precisely 45 minutes on the first Wednesday of every month, preferring to let his managers manage for the rest of the time.

One day Angela, Mr MacCawber's secretary rang Geoff with what she said was some important news. However, Angela told Geoff that she couldn't tell him the news, as that would be a betrayal of confidence, so she would just have to provide him with some clues, and leave him to guess.

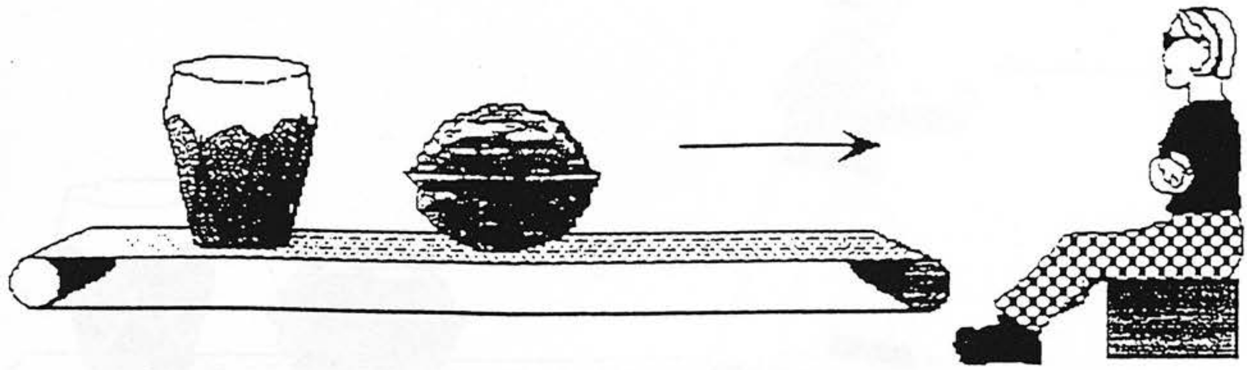
"Well Geoff," said Angela, "Do you remember Birt, who used to run an ice-cream parlour, also called MacCawber's, that was also in the business district of Nawatobee? Last December, Mr MacCawber visited on a day other than usual, going to the ice-cream parlour on the Tuesday, and he sacked Birt for incompetence. Mr MacCawber was heard to say 'I always like to deal with unpleasant business on Mondays and Tuesdays - it gets me in a mean mood for the rest of the week!'

"Secondly, you may be familiar with the case of Marge. Marge manages MacCawber's large car-part dealership - called Part-U-Need - in the suburbs. In January, Mr MacCawber altered his visit to her, going to the dealership on the Thursday instead of the Wednesday. Marge had been given a big raise. Mr MacCawber had been heard to remark I always like to deal with pleasant business on a Thursday or a Friday - it sets up my my mood for the weekend's golfing!"

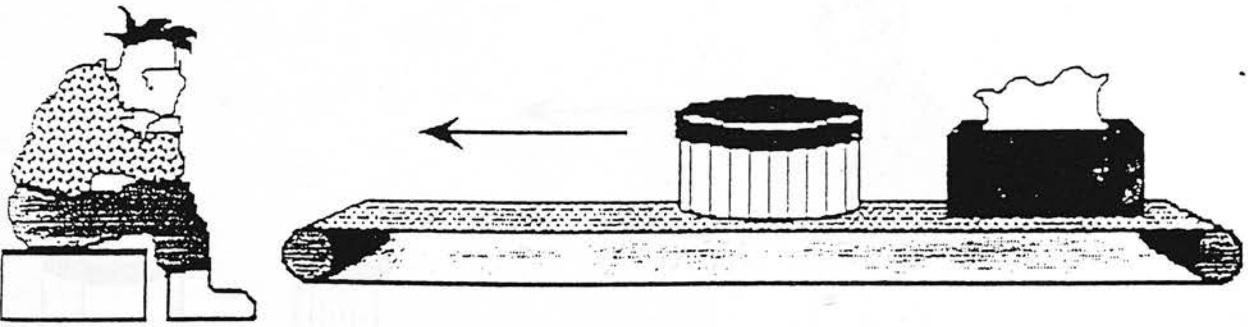
*Object-Moving* Primes (From Boroditsky, 1998):

Please circle TRUE or FALSE for questions 1-3 below.

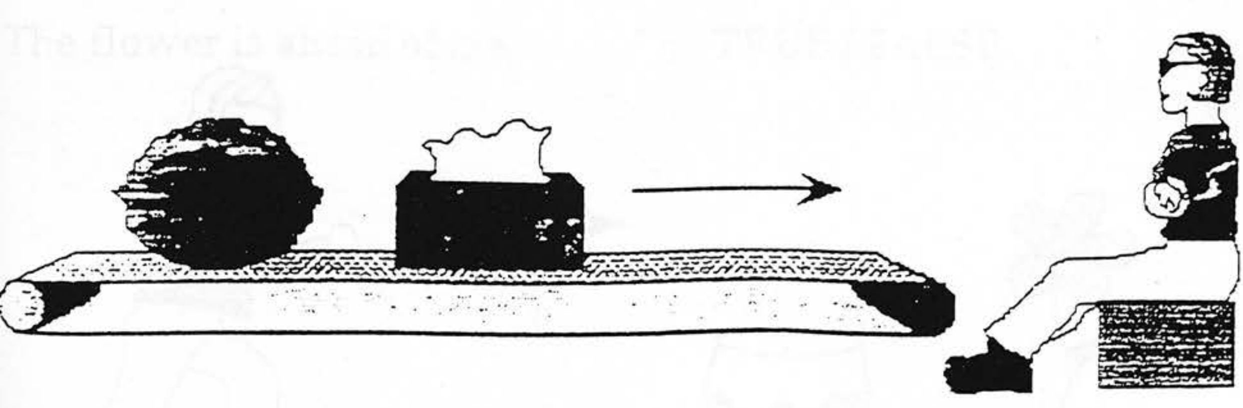
1. The drum is ahead of the walnut. TRUE / FALSE



2. The hat box is ahead of the kleenex. TRUE / FALSE



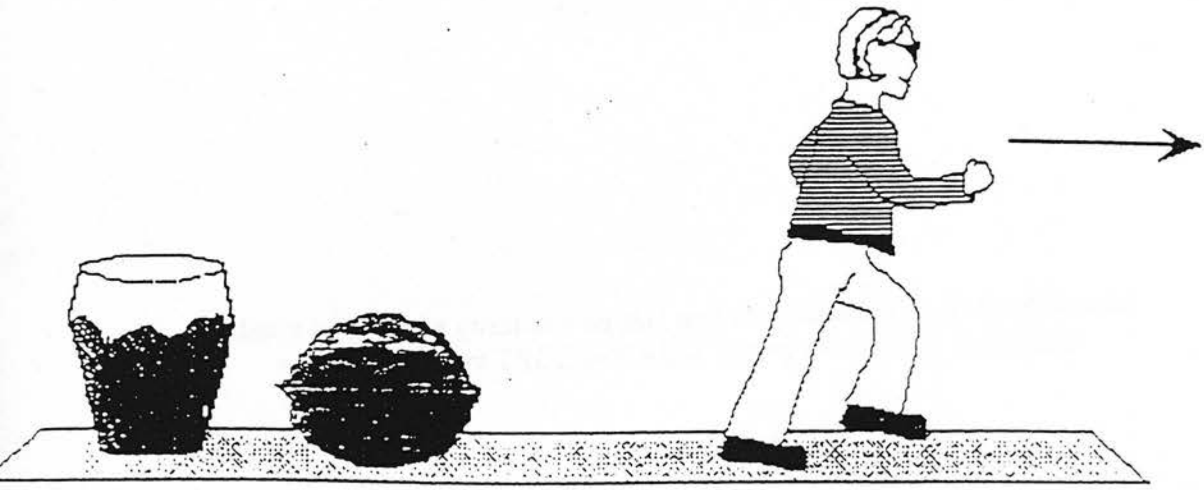
3. The kleenex is ahead of the walnut. TRUE / FALSE



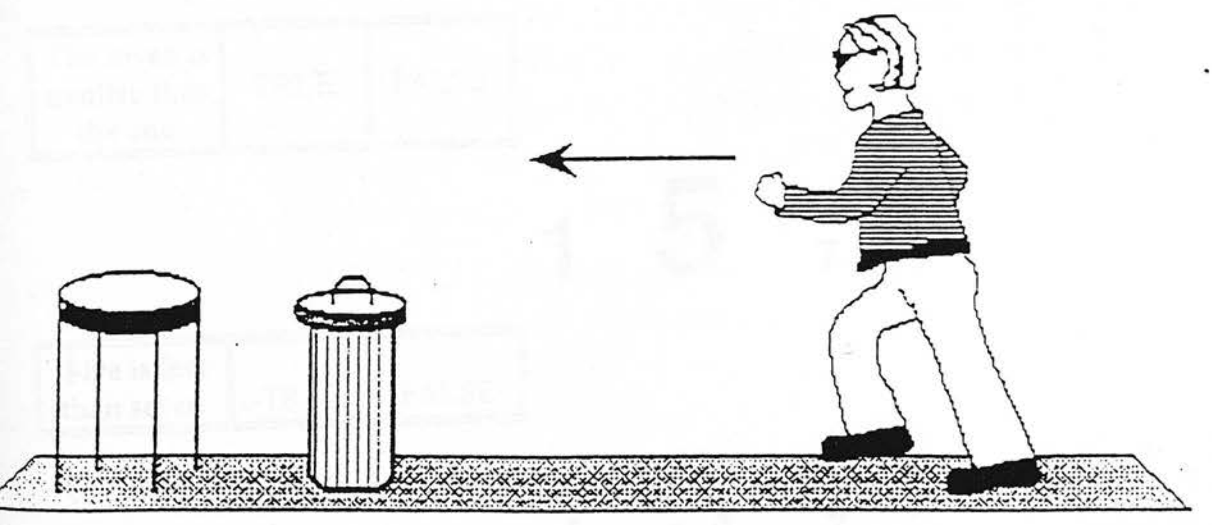
*Ego-Moving* Primes (From Boroditsky, 1998)

Please circle TRUE or FALSE for questions 1-3 below.

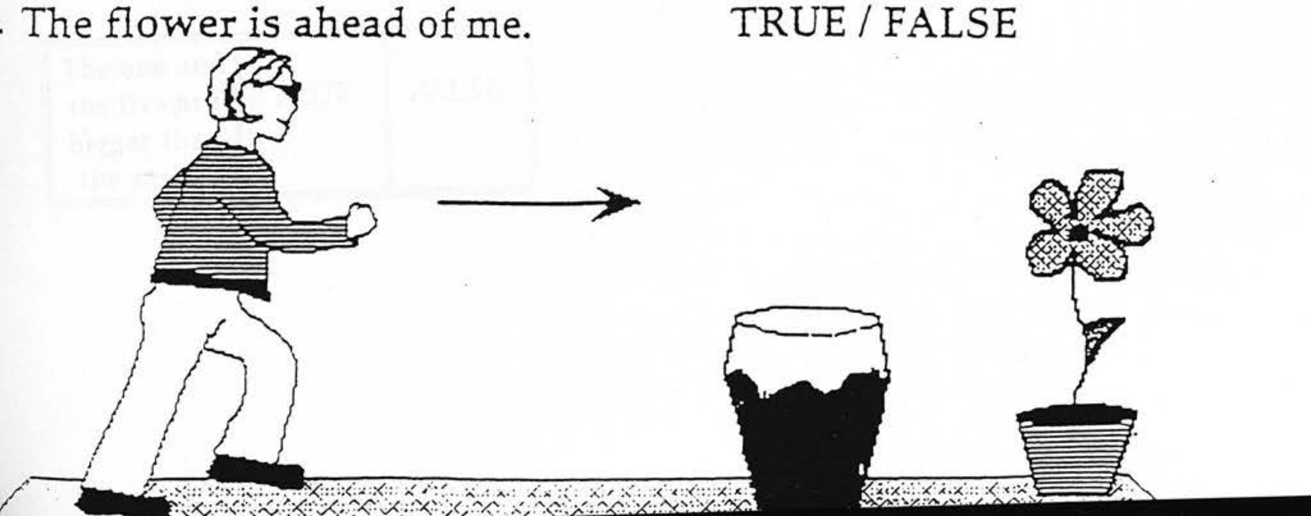
1. The drum is ahead of me. TRUE / FALSE



2. The stool is ahead of me. TRUE / FALSE



3. The flower is ahead of me. TRUE / FALSE



A Sample Non-Priming Divesrion Task.

*Please answer the questions on this and the following page by indicating whether they are TRUE or FALSE (using the buttons provided).*

**1** 5 7

| | | |
|---|---|---|
| The seven is smaller than the one | TRUE | FALSE |

1 **5** 7

| | | |
|---|---|---|
| Five is less than seven | TRUE | FALSE |

1 **5** 7

| | | |
|---|---|---|
| The one and the five are bigger than the seven | TRUE | FALSE |

# *Appendix E: Published Papers*

Some of the work described in this thesis has already been published elsewhere. This appendix reproduces the four existing published papers, with permissions from the publishers.

The four papers are:

Ramscar, M.J.A. and Pain, H.G., (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum Associates, New Jersey, pp 346-351.

Ramscar, M.J.A. Pain, H.G., & Cooper, R (1997) Is there a place for semantic similarity in the analogical mapping process? *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society.* Lawrence Earlbaum Associates., New Jersey, pp 632-636.

Ramscar, M.J.A., Pain, H.G., Darrington, S. and Lee, J. R. (1998) Examples and generalisations: Using surface versus structural recall biases to probe conceptual storage. in Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, Madison-Wisconsin, USA, August, 1998, Laurence Ealbaum Associates, pp 859-865.

Darrington, S., Lingstadt, T and Ramscar, M.J.A. (1998) Analogy as a sub-process of categorisation. in Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, Madison-Wisconsin, USA, August, 1998,Laurence Ealbaum Associates, pp 279-284.

The papers are reproduced in the order shown above.

# Can a real distinction be made between cognitive theories of analogy and categorisation?

**Michael Ramscar† †† and Helen Pain†**
†Department of Artificial Intelligence
††EdCAAD, Department of Architecture
University of Edinburgh
{michael,helen}@aisb.ed.ac.uk

## Abstract

Analogy has traditionally been defined by use of a contrast definition: analogies represent associations or connections between things distinct from the 'normal' associations or connections determined by our 'ordinary' concepts and categories. Research into analogy, however, is also distinct from research into concepts and categories in terms of the richness of its process models. A number of detailed, plausible models of the analogical process exist (Forbus, Gentner and Law, 1995; Holyoak and Thagard, 1995): the same cannot be said of categorisation.

In this paper we argue that in the absence of an acceptable account of categorisation, this contrast definition amounts to little more than a convenient fiction which, whilst useful in constraining the scope of cognitive investigations, confuses the relationship between analogy and categorisation, and prevents models of these processes from informing one another. We present a study which addresses directly the question of whether analogy can be distinguished from categorisation by contrasting categorisational and analogical processes, and following from this, whether theories of analogy, notably Gentner's structure mapping theory (Gentner, 1983; Forbus et al, *ibid.*), can also be used to model parts of the categorisation process.

## Introduction

Ordinarily one accepts a distinction between category membership and analogy according to realist terms. In categorical judgements, relating a new representation of an object to some kind of stored category representation, objects are felt to be similar to one another in a way in which those objects in judgements of analogical association are not. If two objects are considered to be members of a category, the classification is real; if they are considered to be analogous, it is not. Consider, for example an analogy between a theory and a building (Lakoff and Johnson 1980): we might talk of "the *foundations* of a theory"; "we might wish to *buttress* a theory with more facts"; "theories that we *construct* can also *collapse*". From an everyday, psychologically realist viewpoint, an igloo and a castle and a skyscraper really are similar in a way that similarities between buildings and theories are not.

Research into analogy and metaphor has accepted this tacit realism. Holyoak and Thagard (1995) describe a world in which "we think we see things as they really are", and analogy is used in order to recycle our existing knowledge of the real world to formulate new bits of 'real' knowledge. Similarly, in the case of metaphor, Ortony (1979) makes a distinction between literal and non-literal similarities: 'encyclopaedias are like dictionaries' is true in a literal (real) way, whereas 'encyclopaedias are like goldmines' is only true in a metaphorical (non-real) way. Whether the notion of literal similarity might be problematic or not is barely examined, since the real problem to be addressed is metaphor. Holyoak and Thagard (1995) offer the comment "A metaphor always connects two domains in a way that goes beyond our normal category structure" (pp 217), whilst giving little indication as to what might constitute this 'normal category structure'. Analogies are defined as being distinct from categories, the nature of which are left unexamined, presumed real.

Once the difficulties of giving an account of categorisation are admitted into the picture, distinctions between analogy and metaphor reliant upon a contrast with categorisation cease to distinguish at all. Analogy is consistently defined in contrast to categorisation (Clement and Gentner, 1991; Holyoak and Thagard, 1995); yet in order to make a contrast definition one needs an account of at least one of the contrasting elements. This we don't have. An analogy is defined as an associative judgement between two things that are in different categories, yet an account of what constitutes an association between two things such that they are members of the same category rather than different categories is not available (Medin, Goldstone and Gentner, 1993)[1]. Moreover, on the best accounts of categorisation, the question of whether two things are members of the same category may not be amenable to any straightforward answer (Medin and Ortony, 1989; Ramscar, 1996). Thus analogy

[1] Similarly, Glucksberg and Keysar (1990) argue that metaphorical judgements are the *same* as categorisational judgements ("metaphors are understood as they are - as class inclusion statements", pp17). It is hard to see how categorisation is to illuminate metaphor, since they conclude: "The central problem is to understand categorization." ( pp 17).

tends to be defined in contrast to what is in itself a largely undefined process. In the light of this, a definition such as:

"In an analogy, a familiar domain is used to understand a novel domain in order to highlight important similarities between the domains, or to predict new features of the novel domain." *[We interpret domain here to be equivalent to category]* (Clement and Gentner, 1991)

might be more accurately reformulated along the lines of: 'in analogy, a stored representation is used in order to highlight important similarities between it and a new representation of an object or concept, or to predict new features in the new representation of an object or concept'. None of this would be out of place in a definition of categorisation. The distinction between categorisation and analogy is difficult to draw: here we explore the hypothesis that at cognitive levels of description there may no clear distinction to be made at all.

## Models of analogy and categorisation

Another factor which favours the abandoning of traditional distinctions between categorisation and analogy are the strong parallels which can be drawn between theories of analogy and the most plausible models of categorisation. It is becoming more widely accepted that structure plays a major role in category formation (Boyd, 1984; Goldstone, 1994; Kiel, 1989; Medin and Ortony, 1989): analogical reasoning research directly addresses a process which reasons amongst structural networks (Falkenhainer, Forbus and Gentner, 1989; Holyoak and Thagard, 1995). Forbus, Gentner and Law (1995; pp 145-6) propose the following theoretical model of analogical reasoning:

• initial selection dependant upon surface similarity

• analogical similarity is determined by deeper structures

this is strikingly similar to Medin and Ortony's (1989; pp 185-6) knowledge representation scheme for categorisation:

• identification procedure based upon surface features

• classification is determined by deeper structures.

Where research into analogy differs from research into categorisation is in the richness of its process models. A number of detailed, plausible models of the analogical process exist (Forbus, Gentner and Law, 1995; Holyoak and Thagard, 1995): the same cannot be said of categorisation. Medin and Ortony offer little detail as to the mechanisms by which surface identification is governed by deeper structures, or indeed the composition of these deeper structures. In the current study we address directly the question discussed above, of whether analogy can be distinguished from categorisation by contrasting categorisational and analogical processes, and following from this, whether theories of analogy can also be used to model parts of the categorisation process.

## Gentner's structure mapping theory

Gentner's (1983) Structure Mapping Theory is an attempt to explain how it is that two domains can be considered analogous, and in particular how it is that correspondences between analogues from two domains can be mapped. Structure mapping proposes that the mapping and inference between two domains can be achieved by assigning correspondences between objects and attributes and then mapping predicates with identical names. In order to do this, Gentner assumes a predicate like representation distinguishing between *objects*, *object-attributes* and *relations*. Object-attributes are those predicates that have one argument and describe object properties, e.g. `RED(lobster)`. Relations are divided into a hierarchy of orders, with those predicates with two or more arguments which are used to describe relations between objects, for example `UPSETS(stomach,lobster)` forming the lowest order, and those predicates describing different levels of relationships between relations forming the higher orders e.g.: `CAUSE(UPSETS(stomach,lobster),DRINKS(seltzer,diner))`.

The theory itself comprises two parts: *mapping rules*, and the *systematicity principle*. Mapping rules state that (a) attributes of objects are not mapped and (b) relations between objects are preserved. The systematicity principle requires that higher order relations (e.g. `CAUSE` above) are mapped preferentially, followed by the relations that constitute the higher order arguments.

The question of how analogies are accessed, i.e. how representations are selected in order to allow analogical mapping to take place, was addressed experimentally by Gentner, Ratterman and Forbus (1993). Their study showed that analogical access relied primarily upon surface (feature) matches, and they propose that judgements of analogical similarity can be decomposed into two sub-processes:

• *Accessing* a similar (*base*) situation from memory, based primarily on surface similarity

• Creating a *mapping* from base to target using structural commonalities.

## Structural systematicity and categorisation

Since the Gentner, Ratterman and Forbus (1993) studies did not directly address categorisation, a tacitly realist position was adopted in respect of the categories amongst which subjects were to analogise (Ramscar, 1996). The most obvious way in which this realist assumption manifests itself is in the classification of match items (the individual stories within the "Karla the hawk" story sets (Gentner, Ratterman and Forbus 1993)). The question of the categorical status of match items is determined in advance, thus story 1 in figure 1 is classified as a base story, whilst story 3 is defined as its analogue. It is tacitly assumed that the two stories are members of distinct and separate categories, and that they share some kind of analogous link. Whilst the study aimed to explore a wider range of determinants of similarity, the particular correspondences determined by structural systematicity were considered to be indicative of analogous similarities (similarities *between* rather than *within* categories). These assumptions determined the predictions that Gentner *et al* made for their experiments, and the evidence they sought with which to test them.

Gentner *et al*'s study explored criteria of similarity, and discovered that the preferred determinant of analogical similarity in subjects was shared structural systematicity. As a consequence of our hypothesis we predicted that if we were to use Gentner et al's methods and materials to explore categorisation rather than analogy, structural systematicity might also serve as a criterion for the determining category membership. Story 3 in figure 1 was assumed by Gentner et al to be an analogue of story 1. Analogues, as posited in aditional accounts of analogy, are defined in contrast to ategory members. If subjects were to use structural ystematicity as a categorisation determinant, then efinitions of analogy which rely on shared structure to ontrast analogy with categorisation might need some efinement. If both analogy and categorisation produce the ame results, then this might imply some shared, structure ased mechanism, or that one process is supervenient upon ie other. Accordingly, we experimented by presenting ubjects with Gentner *et al*'s materials and asking them to ategorise them. Given that Gentner *et al* define the nalogical mechanism in terms of structure mapping, we ccordingly expected structure mapping to determine ategorisation: i.e. Gentner *et al* assume that match items ith only structural similarities (i.e. analogues) belong to ifferent categories: we predict that they will be categorised gether.

## The Experiment

### ubjects

he subjects were 20 volunteers, a mixture of postgraduate nd undergraduate students from the Artificial Intelligence epartment at the University of Edinburgh.

### laterials

he basic materials used in this study were the 20 sets of Karla the hawk" stories (Gentner, Ratterman and Forbus, 993).[2]

Gentner defines the following taxonomy of similarity elationships between the stories:

- *Literal similarity* matches include both common relational structure and common object descriptions;
- *Surface matches*: based upon common object descriptions, plus some first order relations;
- *Structural similarity*, a match based upon a common system of internal relations;
- *First order matches*, where the only common feature is first order relations;
- *Object only matches*, where stories have only object matches in common.

Each set consists of a base (B), a literally similar story LS), an analogue (TA - with only structural similarities vith the base), a mere-appearance story (MA - with surface nd first order commonalities with the base), a false analogy FA - an analogue of MA), and an object only match story OO - with only surface commonalities with the base). This llowed for a number of potential groupings according to the lassification strategy adopted. Our prediction was that ubjects would use structural similarity as their categorical imilarity determinant, putting analogues and bases into the ame categories (i.e. B, LS and TA together), rather than rouping match items at the object level (i.e. grouping B, S, MA and OO together).

The sets were modified slightly: in Gentner et al's analogy esearch questions of the asymmetry and direction of omparisons were clearly fixed (all comparisons were in relation to the base story). Extra features (a varied mix of objects, attributes and relationships) were added to (or removed from) the base story representations (Figure 1, bold face) which did little to affect analogical similarity judgements. In categorisation judgements, aspects such as symmetry and directionality may be more fluid. As we predicted that structure would be an important determinant of categorical similarity judgements, and noting that the directionality of similarity judgements cannot be fixed in

*Story 1 - Base story*

**Once there was a teacher named Mrs Jackson who wanted a salary increase. One day, the principal said that he was increasing his own salary by 20 percent. However, he said there was not enough money to give the teachers a salary increase.**

When Mrs Jackson heard this she became so angry that she decided to take revenge. The next day, Mrs Jackson used gasoline to set fire to the principal's office.

**Then she went to a bar and got drunk.**

*Story 2 - Literal similarity*

Professor Rosie McGhee very much wanted a raise. One day the provost announced that he was giving himself a raise. However, he said that since money was short, no one else would get a raise this year.

After Professor McGhee heard this she became so upset that she decided to get even. One hour later, Professor McGhee blew up the administration building with dynamite.

*Story 3 - True Analogy*

McGhee was a sailor who wanted a few days of vacation on land. One day, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on the ship.

After McGhee heard this he became so upset that he decided to get revenge. Within an hour McGhee blew up the captain's cabin with dynamite.

*Story 4 - Mere appearance: (First order commonalities)*

Professor McGhee very much wanted a raise. One day she became so impatient that she used kerosene to burn down the administration building.

After the fire, the provost announced that he was giving himself a raise. However, he said that due to the fire, there was not enough money to give one to anyone else.

*Story 5 - False Analogy*

McGhee was a sailor who wanted a few days of vacation on land. One day McGhee became so impatient that he tried to blow up the captain's cabin using dynamite.

After this incident, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on board to repair the ship.

*Story 6 - Mere appearance: (Object commonalities only)*

A teacher once thought that she deserved a pay rise. She asked the principal when her rise was due. She was wearing her best suit. The principal told her that rises were decided by the governors.

**Figure 1:** Sample stories from Gentner, Ratterman and Forbus (1993) - the text in bold type illustrates extra structure added by Gentner at al to the base stories only.

[2] Many thanks to Dedre Gentner for providing the story sets.

categorisation, we accordingly removed Gentner et al's extra features from 65% of the story sets (G- sets: in these, for example, the base / literal similarity relationship were symmetrical), and retained the extra features (and any attendant asymmetries) in 35% of stories (G+ sets).

## Procedure

Subjects were given 10 mixed sets of 6 stories[3] and asked to work through them a set at a time. Both sets and stories were presented in randomised order. For each set, they read through each story a number of times in order to familiarise themselves with its content. Subjects were then asked to "Group the stories into the categories that seemed most natural and appropriate to you. These groups can range from putting every member of the story set into the same group, to putting each story into a group on its own." When subjects had made their categorisation decisions, they physically grouped each set of stories by pasting them onto a large sheet of paper and encircling each group in ink. Subjects were then re-presented with their groupings a set at a time, asked to give each group with two or more members

| Classification Criterion | | | | % of Total |
|---|---|---|---|---|
| *Systematic network of relations in common* - Type 1 | | | | 79.5 % |
| 1 B LS TA | 2 FA MA | 3 OO | | |
| *Systematic network of relations in common* - Type 2 *(Base classified separately)* | | | | 8 % |
| 1 LS TA | 2 FA MA | 3 B | 4 OO | |
| *First order relations in common* - Type 3 | | | | 4 % |
| 1 B LS TA FA MA | 2 OO | | | |
| *Only object similarities in common* - Types 6 & 7 | | | | 5 % |
| 1 MA LS B OO | 2 FA MA | | | |
| 1 B OO | 2 LS MA | 3 TA FA | | |
| *No classification possible* - Types 4, 5, 8 , 9 10 | | | | 3.5 % |
| 1 B LS MA | 2 FA TA | 3. OO | | |
| 1 B TA | 2 FA MA | 3 LS | 4 OO | |
| 1 B MA FA | 2 LS TA | 3 OO | | |
| 1 B LS TA OO | 2 FA MA | | | |
| 1 FA B TA MA | 2 LS | | | |

**Figure 2:** Output patterns from the categorisation task, showing the groups formed and criteria established. The stories are labelled according to Gentner's taxonomy of similarity (defined above): B = Base; LS = Literal Similarity; TA = True Analogy; FA = False Analogy; MA = Mere Appearance; OO = Object Only match.

[3]Given the sample size, we concentrated on sets 1 - 10; sets 11 - 20 were used to a more limited extent to check for any marked variations in the data being produced.

a simple descriptive name, and then to write a few sentences explaining what caused them to classify each named group of stories together[4].

## Results

For each story set the groups formed by each subject's classifications were analysed. The pattern of groupings which emerged fell broadly into 5 types (figure 2). Similarities across groupings (i.e. similarity shared by every member of a two or more member group across a categorised story set) according to Gentner et al's taxonomy of similarities could be identified in 96.5% of groupings. Of these, in 5% of cases the stories were grouped according to types 6 and 7. The only similarities across groupings in these types are that the stories in the individual groups had only objects in common. In 4% the stories were classified according to type 3, where the across grouping similarity was shared first order relations. In 79.5% of cases subjects grouped using type 1. Here the only similarity across groupings was a network of systematic causal relations. The full output and incidence of the types is given in Table 1.

| | Story Set Type | | |
|---|---|---|---|
| Grouping Type | G+ | G- | *% of total* |
| *Type 1* | 68 % | 86.5% | 79.5% |
| *Type 2* | 20 % | 0.5% | 8% |
| *Other* | 12% | 13% | 12.5% |

**Figure 3:** Classification strategies according to set type.

8% of groupings were according to type 2, where the base was put into a category on its own, with the only similarity across other groupings being shared structure. This type was only found once amongst those sets from which Gentner *et al*'s extra features had been removed (0.5% of G- sets; figure 3). The G+ sets, those with added features in the base, were sets 5; 7; 10; 12; 15; 17; and 20. Of these: in set 5 and set 20 the extra features involved higher order relations; in sets 7 10, 15 they involved first order relations; and in sets 12 17 the extra features were objects. 20% of these sets were classified as type 2, with the bulk of these classifications being in the sets with extra higher order relations (figure 4).

| | *Higher-order relations Sets* | | *Objects only Sets* | | *1st order relations Sets* | | |
|---|---|---|---|---|---|---|---|
| | 5 | 20 | 12 | 17 | 7 | 10 | 15 |
| Type 1 | 7 | 2 | 7 | 7 | 11 | 9 | 5 |
| Type 2 | 9 | 2 | 1 | | 2 | | |
| Type 3 | 1 | 1 | | | 1 | 2 | |
| Types 6 & 7 | 2 | | | | | | |
| Types 4, 5, 8, 9 & 10 | | | | | | | 1 |
| Totals | 19 | 5 | 8 | 7 | 14 | 11 | 6 |

**Figure 4:** Classification data for the G+ sets.

[4]This data is currently being analysed, and will not be considered here.

Story Sets

| Subjects | 1 | 2 | 3 | 4 | 5+ | 6 | 7+ | 8 | 9 | 10+ | 11 | 12+ | 13 | 14 | 15+ | 16 | 17+ | 18 | 19 | 20+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | | | | | 1 | 1 | | 1 | 1 | 1 | | | 1 | 1 | 1 | | 1 | 1 | | |
| B | 3 | 3 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | |
| C | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | |
| D | 1 | 1 | 1 | 1 | 2 | | 1 | | | | 1 | 2 | | | | | | | 1 | 2 |
| E | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | | | | | | | | | | |
| F | 4 | 1 | 1 | 1 | 2 | 1 | 1 | | | | 1 | 1 | | | | | | 5 | | |
| G | | 8 | | | 6 | 7 | | 1 | 1 | | | | 1 | 6 | 1 | | 1 | 1 | | |
| H | 5 | 1 | 1 | 1 | 2 | | 1 | | | | 1 | 1 | | | | | | | 1 | 1 |
| I | 1 | 1 | 1 | 1 | 1 | | 1 | | | | 1 | 1 | | | | | | | 1 | 2 |
| J | 1 | | 1 | | 8 | 1 | | 1 | 1 | | 1 | | | | 1 | | 1 | 1 | | |
| K | | | | | | | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | | | 1 |
| L | 7 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | |
| M | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | 1 | 1 | | | | | | 1 | | |
| N | 1 | 1 | 1 | 1 | 5 | 2 | 2 | | | | 1 | 1 | | | | | | 1 | | |
| O | | | | | 1 | 1 | | 1 | 1 | 1 | | | 1 | 1 | 1 | | 1 | | 1 | |
| P | 6 | 3 | 1 | 1 | 1 | | 3 | | | | 7 | 1 | | | | | | | 1 | 3 |
| Q | | | | | 2 | 1 | | 1 | 1 | 1 | | | 2 | 1 | 1 | | 1 | 1 | | |
| R | 1 | 10 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | |
| S | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | |
| T | | | | | 3 | 7 | | 9 | 1 | 3 | | | 7 | | 7 | 7 | 1 | 1 | | |

**Table 1:** Output incidence of subject groupings. Each subject was given 10 story sets (each row represents one subject): the type of grouping is indicated by the type number in the story set column (see also figure 2).
Subject T produced some rather strange results: this was explained by examining the reasons T gave for her groupings, in which she explained that she was exploring a different heuristic for each story set.

## Discussion

Our study explored the hypothesis that mechanisms normally considered to be analogical could in fact support categorisation tasks. The most important finding here is the role that shared structure plays in categorisation judgements. 79.5% of the groupings formed by our subjects had only shared systematic structure (traditionally defined as analogy) as a common feature amongst members of the categories formed. In contrast, only 5% of groupings produced had common object descriptions as the common similarity across categories. Traditionally categorisation models have concentrated on object descriptions, making use of very representationally-simple attribute-value lists (Murphy and Medin, 1985), whereas analogy research has examined relationships between highly structured representations (considering the influence of attributes, relations and higher-order relations in judgements of similarity).

The argument for abandoning the current de facto distinction between categorical and analogical associations of objects is twofold: firstly, that the standard distinctions (Clement and Gentner, 1991; Holyoak and Thagard, 1995) between analogy and categorisation actually failed to distinguish between them; and secondly, that by removing the distinction, understanding of the factors which govern mappings between representations that have been gleaned from analogy research might help illuminate categorisation questions (Ramscar, Lee and Pain, *in press*). Our results provide evidence that structures, and more pertinently Gentner's structural systematicity, rather than features, are the key to categorical similarity in this instance: this tallies with other evidence, such as Rips (1989) who found that subjects were reluctant to change classifications as a result of feature changes alone.

Our argument is supported not only by the proportion of categorisations that were determined by commonalities between internal structures in the stories, but also by the effects of added structure in the G+ sets where the added structure was a higher-order structure. These might at first appear to present a problem for our attempt to use a structure mapping analysis to model these categorisation judgements. In these cases, Gentner's base stories were put into separate categories from stories to which they were supposed to be literally similar, which were in turn categorised alongside their supposed analogues (both of which were supposed to share structures with the base).

These results can be attributed to the effects of directionality and symmetry upon similarity judgements. Whilst Gentner, Ratterman and Forbus (1993) found that subjects judged literally similar (LS) stories to be very similar to bases, and analogues less so, they did not consider the effect of reversing the directionality and symmetry of the comparisons, for example comparing the base and analogue stories similarity to the LS. Neither did they consider the judging of cumulative similarity, where dissimilarities are also taken into account. During this process, the structural dissimilarities of the base versus the LS and analogue appear from our results to be clearly relevant, whereas the object differences of the analogue versus the LS and base do not. This maximisation of important similarities (i.e. structure matches) relative to lesser dissimilarities (i.e. object matches) amongst groupings appears to play a crucial role in categorisation in this study. Whilst it might be argued that all we have shown here is that subjects will form categories of analogies, such an interpretation (in so far as we can make sense of it) does not affect our argument that it is structure that determines the content of these categories.

All of this strengthens our dubiety with respect to the separation of analogy from categorisation. We should note, however, that asserting that analogy cannot be distinguished from categorisation at a cognitive level is not the same thing as arguing that analogy is the same thing as categorisation. Categorisation is such a central cognitive process that it is hard to see how it can be reduced to a single process (c.f. Goldstone, 1994). It may well be that any given manifest reasoning process - such as rule following or metaphor - might be able to illuminate some aspect of categorisation: i.e. can provide the constraints necessary to determining certain categorical similarities. We argue that the analogical process cannot be distinguished from the 'categorisation process' at a cognitive level. Our hypothesis is that analogy is supervenient upon an important part of the classification process, and that as such analogy research is capable of illuminating[5] some categorisation tasks, for instance, the way in which structural systematicity can determine both analogical and category judgements.

## Acknowledgements

Many thanks to Paul Brna, John Lee and Richard White for their contribution to this work, and to Shari Trewin and Robin James for comments on earlier drafts of this paper.

We would also like to acknowledge the helpful and useful comments provided by the anonymous reviewers of this paper.

This work was funded in part by EPSRC Grant GR/J76897.

[5] For example, one problem faced by all cognitive theories of categorisation is explaining typicality effects (e.g. Rosch, 1978); how they occur, or even, how the existence of typicality effects can be accommodated by a given model. Gentner et al (1993) have shown how differing aspects of similarity - structural versus surface - affect recall, soundness ratings and judgements of similarity. By showing that judgements of categorical similarity and the recall of category members can be reliant upon different representational features (surface attributes for recall, structural systematicity for similarity and typicality), we might be able to begin to present a model of the categorisation process which can explain and account for at least some typicality effects.

## References

Boyd, R. (1984) Natural kinds, homeostasis, and the limits of essentialism. Paper presented at Cornell University

Falkenhainer, B., Forbus, K.D. and Gentner, D. (1989) The structure mapping engine: an algorithm and examples. *Artificial Intelligence*, **41**, 1-63

Clement C.A, and Gentner, D. (1991) Systematicity as a selection constraint in analogical mapping *Cognitive Science*, 15: 89-132.

Gentner, D. (1983) Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, **7**: 155-170.

Forbus K, Gentner, D, and Law , K (1995) MAC/FAC: A model of similarity based retrieval *Cognitive Science* **19**:2:141-205

Gentner, D Ratterman, M.J. and Forbus, K. (1993) The roles of similarity in transfer. *Cognitive Psychology* **25**: 524-575

Goldstone, R.L. (1994) The role of similarity in categorization: providing a groundwork. *Cognition*, **52**: 125-157

Glucksberg, S. and Keysar, B. (1990) Understanding metaphorical comparisons: Beyond Similarity. *Psychological Review*, **97**: 1: 3-18

Holyoak, K.J. and Thagard, P. (1995) *Mental Leaps*. MIT Press, Cambridge, Ma.

Kiel, F.C. (1989) *Concepts, kinds and cognitive development*. MIT Press, Cambridge, Mass.

Lakoff, G and Johnson, M (1980) *Metaphors we live by*. University of Chicago Press, Chicago, Ill.

Medin, D and Ortony, A. (1989) What is psychological essentialism? In S. Vosniardou and A. Ortony (Eds) *Similarity and analogical reasoning*. Cambridge University Press.

Medin, D., Goldstone, R. and Gentner, D. (1993) Respects for similarity. *Psychological review*, **100**:2:254-278.

Murphy, G.L. and Medin, D.L. (1985) The role of theories in conceptual coherence. *Psychological review*, **92**, 289 - 316

Ortony, A. (1979) The role of similarity in similes and metaphor. In Ortony, A. (Ed) *Metaphor and thought*. Cambridge University Press.

Ramscar M.J.A. (1996) Judgements of association: connecting cognitive theories of analogy and categorisation - DAI Research Report, Department of Artificial Intelligence, University of Edinburgh, Scotland.

Ramscar, M. J. A, Lee, J. R. and Pain, H (*in press*) A classification-based methodology for the integration of agent views within design systems. To appear in *Design Studies*, Vol. 17, Autumn 1996 .

Rips, L.J. (1989) Similarity, typicality and Categorization. In S. Vosniardou and A. Ortony (Eds) *Similarity and analogical reasoning*. Cambridge University Press.

Rosch, E. (1978) *Cognition and Categorisation*. Lawrence Earlbaum Assoc., New Jersey.

# Is there a Place for Semantic Similarity in the Analogical Mapping Process?


**Michael Ramscar and Helen Pain**
Department of Artificial Intelligence
University of Edinburgh
{michael,helen}@aisb.ed.ac.uk

**Richard Cooper**
Department of Psychology
Birkbeck College, University of London
R.Cooper@psyc.bbk.ac.uk



## Abstract

Ramscar & Pain (1996) argued that the analogical process cannot be easily distinguished from the categorisation process at a cognitive level. In light of the absence of any distinction between analogy and categorisation, we have argued that analogy is supervenient upon an important part of the classification process, and that as such 'analogical' models are capable of illuminating some categorisation tasks, for instance, the way in which structural systematicity can determine not only analogical judgements, but also category decisions. Our scepticism regarding the cognitive distinction between these two processes has implications for both analogy and categorisation research: in this paper we consider two leading analogical theories, Gentner's Structure Mapping Theory and Holyoak's Multi-Constraint Theory, and argue that results from our use of analogical modeling techniques in categorisation tasks offer some important insights into exactly which elements should be included in a theory of analogical mapping.


## Introduction

Ramscar & Pain (1996) argue for a position that rejects any simple distinction between analogy and categorisation at the cognitive process level (see Ramscar, Pain & Lee, 1997 for more evidence). The argument runs as follows: definitions of analogy (e.g. Holyoak & Thagard, 1995; Clement & Gentner, 1991) rely on making a distinction between 'straight' categorical judgements on the one hand, and analogical, 'extra-categorical' judgements on the other; with no convincing account of just what constitutes a 'straight' categorical judgment, this amounts to little more than hand waving, attempting to characterise one ill-defined process by contrasting it with another ill-defined process. In the absence of any compelling reason to believe in distinct cognitive processes of analogy and categorisation, there is much utility in viewing human analogical and categorical behaviour as manifestations of the same process: research into analogy has yielded a number of plausible models of the analogical process (Forbus, Gentner & Law 1995; Holyoak & Thagard, 1995), in sharp contrast to categorisation research, where process models, to the extent that they feature at all, tend to be more conjectural in nature (e.g. Medin & Ortony, 1989). A companion paper to this (Ramscar, Pain & Lee, 1997) presents further evidence of the benefits and insights that this approach can bring to investigations into the nature of categorisation decisions.

In this discussion, we wish to examine the perspective our particular view of the categorisation / analogy divide, and the results of our experiments using analogical modeling to explore categorisation, can bring to theoretical approaches to analogy (and by implication, existing models of analogy). The removal of neatly bounded 'concept domains' with which various aspects of the analogical process can interact produces a changed circumstance that seems likely to have repercussions for theories based on just such assumptions: in particular, we wish to examine the implications of our approach to the kinds of similarity constraints — structural, pragmatic and semantic — which are variously included in, or excluded from, theories and models of the analogical process.

A number of 'competing' theories of analogy exist: Holyoak & Thagard (1995); Gentner (1983; Forbus, Gentner & Law, 1995); Keane (Keane, Ledgeway & Duff, 1994); Hofstadter (1995). In this paper we examine in detail the effects of our blurred distinction perspective on the vexed question of 'semantics' in the first two of these, Gentner's 'Structure Mapping' theory and Holyoak & Thagard's 'Multi-Constraint' theory. These constitute the most explicit theories, whose supporting evidence and accompanying process models have been widely disseminated and accepted.

## Mapping Without The Distinction

### Gentner's Structure Mapping Theory

Gentner (Gentner, 1983; Clement & Gentner, 1991) proposed the Structure Mapping Theory as an attempt to explain how it is that two domains can be considered analogous, and in particular how it is that correspondences between analogues from two domains can be mapped.

Structure mapping proposes that the mapping and inference between two domains can be achieved by assigning correspondences between objects and attributes and then mapping predicates with identical names. In order to do this, Gentner assumes a predicate like representation (figure 1), distinguishing between *objects, object-attributes* and *relations*. Object-attributes are those predicates that have one argument and describe object properties, e.g. YELLOW(sun). Relations are divided into a hierarchy of orders, with those predicates with two or more arguments which are used to describe relations between objects, for example ATTRACTS(sun, planet) forming the lowest order, and those predicates describing different levels of relationships between relations e.g. CAUSE(ATTRACTS(sun, planet), REVOLVES_AROUND(planet, sun)) forming the higher orders.

*Solar System*

YELLOW(sun)
MASSIVE (sun)
HOT(sun)
ATTRACTS(sun, planet)
ATTRACTS(planet, sun)
MORE_MASSIVE_THAN(sun, planet)
REVOLVES_AROUND(planet, sun)
HOTTER_THAN(sun, planet)

*Hydrogen Atom*

ATTRACTS(nucleus, electron)
ATTRACTS(electron, nucleus)
MORE_MASSIVE_THAN(nucleus, electron)
REVOLVES_AROUND(electron, nucleus)

Figure 1: Predicate representations of the solar system and a hydrogen atom

The theory itself comprises two parts: *mapping rules*, and the *systematicity principle*. Mapping rules state that attributes of objects are not mapped; but relations between objects are preserved. The systematicity principle requires that complex higher order relations (e.g. CAUSE above) are mapped preferentially, followed by relations that constitute the higher order arguments. This is intended to capture the notion that analogy conveys a system of connected knowledge, rather than an assortment of independent facts:

> "structure mapping stems in part from the observation that useful analogies, such as those used in science or education, involve rich, interconstraining systems of mappings between two domains, rather than a set of independent correspondences" (Clement & Gentner, 1991, pp. 91–92)

Gentner's structure mapping theory has been implemented as a computer simulation model SME (Falkenhainer, Forbus & Gentner, 1989).

One criticism that has been made of structure mapping theory is that the representations it uses are arbitrary. The primitives used in the examples are chosen selectively. The example would fail if REVOLVES_AROUND in the Solar System domain was replaced with ORBITS. Furthermore, the inclusion of a larger range of domain relations would result in the matches appearing far less sound than they do. Moreover, Holyoak (1985) claims that Gentner can give no account as to why HOTTER_THAN in figure 1 should not be mapped onto the Hydrogen Atom domain. Holyoak argues that expanding the representation in order to utilise the systematicity principle will be of little help, since on another expansion the relation HOTTER_THAN will form part of a higher order system (involving the warming of planets, conditions for the existence of life, etc.). The structure mapping theory seems to give no explanation as to why this relation is not mapped.

To some researchers this has appeared to be a major problem. A system which uses no goal or domain information, and which relies purely on structural inferences, might appear to have trouble when, given a number of networks of higher-order relations between domains (i.e. several causal networks), giving an account as to why it selected the one that was relevant to making the desired analogy.

## Holyoak and Thagard's Multi-Constraint Theory

Holyoak's theory (Holyoak, 1985; Holyoak & Thagard, 1989; Holyoak & Thagard, 1995) is concerned primarily with problem solving. It attempts to capture the intuition held by some researchers that goals play some part in the analogical process, particularly in mapping. (Gentner's theory precludes any such influence.) The scope and ambition of this theory has changed considerably over time: what is presented here is a sketch of its general nature.

Holyoak *et al* advocate an explanation of analogy in terms of a goal-driven processing system — mappings are controlled by the system's goals. The analogical mapping problem is seen as one of explaining how the large number of possible mappings between domains can be evaluated and a subset of these used for the transfer of information between domains.

They suggest that this subset emerges from an attempt to balance the different influences upon the mapping process. More specifically they regard the process as an attempt to simultaneously satisfy several constraints. The first group of these (logical compatibility, role identity, uniqueness and relational consistency) are structural, and therefore compatible with the overall thrust of Gentner's theory. However, Holyoak *et al* also consider constraints of *pragmatic centrality* and *semantic similarity* to be integral to the mapping process. These non-structural constraints rely on information other than that found in the basic domain representations.

**Structural Constraints** Logical compatibility ensures that mappings are only considered if they are between entities of the same 'type'. Thus, in the Solar System / Hydrogen Atom analogy the predicate MORE_MASSIVE_THAN cannot be matched with the object electron. Similarly, a mapping between the predicates HOT and REVOLVES_AROUND will not be considered as they take different numbers of arguments. This primarily syntactic constraint is intended to ensure that mappings between different levels of description are not attempted. For example single-argument predicates, such as HOT, tend to be purely descriptive, specifying a particular attribute of an object. Multi-argument predicates describe relationships between objects, and so can be considered to represent a higher level of description. The argument is that mappings between different levels of description are not productive and this constraint serves to eliminate any potential mappings of this kind.

A further hurdle potential mappings must overcome in this model is the role identity constraint. This assumes that the base and target domains can be divided at a higher level of description than that at which the mapping takes place. In the use of analogy in problem solving, upon which the authors focus predominantly, this means the domains may be redescribed in terms of a start state, the problem goals, and the operators that can be used to try and achieve these goals. Role identity then limits mapping to relations and objects that appear in the same part of the domain definition. This provides a weak pragmatic influence in that elements can only be considered for mapping if they play a similar role in both domains.

Holyoak *et al* assume that each element in the base domain will ultimately map onto only one element in the target domain, and *vice versa*. Thus there is a competition between members of the set of potential mappings between one base element and a number of possible target elements. For example, if HOTTER_THAN in the base maps onto HOTTER_THAN in the target then it cannot map onto LESS_MASSIVE_THAN in the target. Accordingly any factor which serves to increase the level of support for the former mapping will consequently act to decrease support for the latter.

A final structural constraint is relational consistency, which acts to ensure that mappings between the base and target domains are consistent. If mappings between structural elements receive support, mappings between the structures themselves, and any other elements, are also supported.

**Pragmatic Centrality** The importance of an element (object or relation), whether in the base or target domain, is another consideration in the mapping process. An element's importance is defined in terms of how useful the element is in satisfying the current goal (or subgoal) of the 'analogiser'. Thus any mappings involving 'useful' elements receive more support than mappings involving less useful elements. When our example analogy is used to explain the relative motion of sub-atomic particles, mappings involving YELLOW and HOTTER_THAN are going to be less favourably considered than those involving REVOLVES_AROUND, since the former are not utilised in satisfying any explanatory goals.

**Semantic Similarity** Holyoak *et al* suggest that the most useful mappings are likely to come from elements which are semantically similar. In the Solar System / Hydrogen Atom analogy, predicates with identical names can be regarded as more similar than those with different names. In more complex examples the method of determining relative similarity is more difficult. Holyoak *et al* make no claim as to any particular model of semantics. The semantic similarity constraint is regarded more as a heuristic than a firm rule, and can be applied in differing strengths at various stages of the mapping process.

Holyoak & Thagard (1989) emphasise that the logical compatibility and role identity constraints are restrictions on the building of the mapping network, and these restrictions are regarded as less important than the three principal constraints of isomorphism (uniqueness and relational consistency), semantic similarity and pragmatic centrality.

The theory has been implemented as a computer simulation ACME, which is a constraint satisfaction network incorporation these considerations.

## Comparing The Two Theories

**Structure** Holyoak & Thagard (1995) argue that the similarities and differences between their theory and structure mapping theory can best be illustrated by comparing Gentner's theory with the three main constraints posited by multi-constraint theory. They claim that multi-constraint theory captures Gentner's insight regarding the importance of systematic structure (in ACME, interconnected systems will have more mutually supporting links than an isolated relation), but in a more flexible manner. As implemented in SME, Gentner's theory rigidly enforces one-to-one mappings and structural consistency — potential mappings which violate these constraints are not made. In contrast, ACME, whilst preferring one-to-one mappings (by using inhibitory links to discourage many-to-one mappings) nevertheless will allow violations.

**Pragmatic Constraints** Gentner's theory (and SME) does not incorporate or recognise the influence of pragmatic — goal driven — constraints on the mapping process. According to structure mapping theory, the operation of goals is external to the actual mapping process, constraining the evaluation of mapping outcomes, rather than actual mappings (although I-SME (Forbus, Ferguson & Gentner, 1994) does incorporate pragmatic influences in mapping). However, whilst the evidence that goals can influence *what is mapped* in analogy is clear (Spellman & Holyoak, 1992), it is less clear that goals directly influence or constrain the *mapping process*.

We noted two criticism of structure mapping earlier: the neatness of the representations used in SME (this applies equally to ACME), and that if an analog offers up two competing possible modes of transfer with a similar level of systematicity, then the systematicity principle cannot act as a constraint in the selection of one or the other. None of this need militate against structure mapping in principle: an analogue which allows two equal mappings may be a poor choice of an analogue; the systematicity principle's yielding of two equally valid mappings may be a psychologically valid resolution of the initially poor choice of analogy. To return to the earlier objection to Gentner's structure mapping theory (Holyoak, 1985), it might well turn out that an expansion of the representations of the Solar System and the Hydrogen Atom will simply lead to a situation in which the two examples are no longer seen to be analogous. Given that the success of any analogy is contingent upon the way in which the putative analogues are represented (Gentner, 1989), it is not a failing of a theory that it cannot provide accurate mappings in situations where candidate analogues are presented in such a way as to obscure any analogous similarity between them.

A major criticism of all analogical theories is that they do not consider the considerable psychological evidence that the choice of representation is crucial to analogy (Hofstadter, 1995). If categorisation research tends to ignore processes and overconcentrate on representation, the opposite is true of analogy. Yet, in fact some of the best evidence of the influence of goals (Spellman & Holyoak, 1992) seems to support the view that goals influence representation rather than mapping.

These "results suggest one way in which analogies can be used systematically to influence people's inferences — [the representation of] the source [analogue] can itself be massaged to encourage a desired mapping" (Holyoak & Thagard, 1995).

**Semantic Constraints** With regard to the semantic similarity constraint, the main focus of this discussion, the respective positions are as follows: SME only matches predicates with identical names, thus if planets (see figure 1) were represented as SMALLER_THAN the sun, and electrons as LESS_MASS_THAN a nucleus, then structure mapping would not allow, and SME would not make, a mapping between the two relations. On the other hand, whilst ACME again prefers to map identical relations, weights on the network can be adjusted to capture the semantic similarity between SMALLER_THAN and LESS_MASS_THAN. Holyoak & Thagard argue that this shows a significant weakness in Gentner's theory:

> "with its emphasis on structure to the exclusion of all other constraints, SME does not simply discourage mappings between non-identical but semantically similar items; it does not even permit them." (Holyoak & Thagard, 1995, p. 258)

Holyoak and Thagard's criticism of the lack of semantic considerations in Gentner's theory carries a lot of intuitive weight. It does seem a perverse, restrictive analogical theory that rejects mappings between SMALLER_THAN and LESS_MASS_THAN in the course of an analogical mapping. However, it is not necessarily so, and from the perspective of a blurred distinction between analogy and categorisation, it might actually be that rejecting such a mapping is necessary *per se*, rather than necessarily perverse.

One reason for so-arguing stems from the results of our investigations into the effects of systematic structure upon categorisation judgements. Ramscar & Pain (1996) addressed the question of whether analogy can be distinguished from categorisation by contrasting categorisational and analogical processes by presenting subjects with Gentner *et al*'s analogy materials and asking them to categorise them. Given that Gentner *et al* define the analogical mechanism in terms of structure mapping, and given our hypothesis that this process was not distinct from a basic categorisation process, we expected structure mapping to determine categorisation. Gentner *et al* seem to implicitly assume that match items with only structural similarities (i.e. analogues) belong to different categories. We predicted that they would be categorised together. We found that 79.5% of the groupings formed by our subjects had only shared systematic structure (traditionally defined as analogy) as a common feature amongst members of the categories formed. In contrast, only 5% of groupings produced had common object descriptions as the common similarity across categories (i.e. the attribute matches often thought to be determinate of categorisation). To the 79.5% of structural congruity groupings could be added a further 8% of classifications where structural *additions* to otherwise structurally congruent representations caused them to be classed singularly. Thus we concluded that that mechanisms normally considered to be analogical could also in fact support categorisation tasks, and in this instance, no discernible difference could be found between analogical and categorical behaviour.

This suggests that the process of classifying two terms together (mapping SMALLER_THAN to LESS_MASS_THAN) is no different to the process of determining the analogy between SOLAR_SYSTEM and HYDROGEN_ATOM. Thus the process of mapping SMALLER_THAN to LESS_MASS_THAN seems to be less a sub-process of the process of determining the analogy between SOLAR_SYSTEM and HYDROGEN_ATOM and more like the same process functioning in parallel.

**Prototype Schemas?** One possible objection to the above would be that our characterisation of categorisation is incomplete. It might be argued that analogical judgements may not be easily distinguished from classification judgements, but *categorisation* judgements can be. This claim relies on the idea that categorisation doesn't involve that same mapping process because it makes use of generalized 'prototypical' schemas (Holyoak & Thagard advocate such a view of categories in Holland *et al*, 1984). If SMALLER_THAN and LESS_MASS_THAN share the same prototypical schema, then there will be no need to compute the similarity between SMALLER_THAN and LESS_MASS_THAN, since such similarity can be confirmed merely by reference to the prototype[1]; and it is this confirmatory reference-to-prototype that is modeled by the semantic links in Holyoak and Thagard's theory.

There are a number of objections to such an account. Practically, there is the problem of providing a convincing account of what a prototype category is (Medin & Ortony, 1989; Ramscar & Pain, 1996). Propositional schematic models of representation (some variety of which is assumed in the majority of cognitive theories) have the power to store a combination of both schemas and the exemplars from which such schemas are constructed. Moreover, what is actually stored is the subject of much debate: are just schemas stored? Fodor & Lepore (1996) give many good reasons for doubting such a 'pure' prototype schema theory. If both are stored, then how do exemplars contribute to schemas? Some have advocated a rejection of schemas altogether (e.g. Nosofsky, 1988), arguing that only exemplars are stored, and that new objects are classified by comparison with stored exemplars, and the calculation of some kind of fit. A further experiment (Ramscar, Pain & Lee, 1997) was designed to see whether analogical theory could shed any light on the nature of subjects' stored representations.

In their analogical recall experiments, Gentner *et al* (Gentner, Ratterman & Forbus, 1993) showed that analogical access relied primarily upon surface attribute (object) matches:

[1] Curiously, a footnote to Gentner, Ratterman & Forbus (1993) states "SME's constraint of matching identical predicates assumes canonical *conceptual* representations, not lexical strings. Two concepts that are similar but not identical (such as "bestow" and "bequeath") are assumed to be decomposed into a canonical representation language so that their similarity is expressed as a partial identity (... "give")", which suggests that this idea has widespread appeal; a version of it has been advocated by the authors in the past (Ramscar, Lee & Pain, 1996).

stories with shared attributes were recalled from memory far more readily than objects with shared structure, even though subjects adjudged some of these stories to be less similar, and inferences generated from them to be less sound with regards to a base than the stories which possessed only shared structure. We decided that this phenomenon might be useful in exploring representation. In the course of the classification experiment, subjects were asked to give their classes "a simple descriptive name", which they could then associate with their classification. By examining what attributes they could recall that were associated with that name, we used Gentner *et al*'s findings about structure versus attribute to shed some light on the mental representation associated with the name. If subjects stored some kind of abstracted prototype, then our hypothesis was that attributes associated with the most prototypical stories would be most readily retrieved from memory, with other attributes recalled insofar as they were shared with the prototypical story. On the other hand, if subjects stored only exemplars, then the lack of context provided by a simple class name should make it equally likely that attributes associated with any given (or all) exemplar might be recalled.

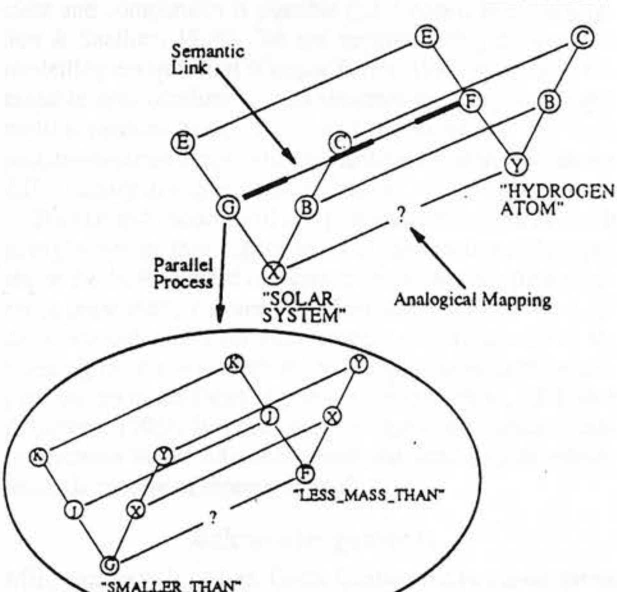



Figure 2: Two views of the mapping process

The experiment yielded little evidence to support the hypothesis that subjects had abstracted and stored schemas from the groups they had classified, despite the fact that it appeared to be a shared structural schema that was the basis of subjects' original classification decisions (Ramscar & Pain, 1996). Instead, when Gentner, Ratterman & Forbus (1993)'s analysis was applied to subjects' behaviour, it appeared that being presented with a class name in no particular context caused the subjects to randomly recall one of the exemplars associated with that name, and then use that exemplar as the stimulus for recalling other class members; overall, only 10% of recall tasks favour the former analysis, wheras some 71% support the latter. Thus as well as finding no evidence to support stored prototypes, Ramscar, Pain & Lee (1997) provide evidence that only exemplars (instances) of categories are stored (which strengthens the argument for not strongly distinguishing analogy and categorisation). It would seem that schemas can't supply the necessary theoretical justification for semantic links.

## Discussion

If we consider the schematic representation of the mapping process in figure 2, Holyoak & Thagard model the mapping of F onto G in analogy $X \rightarrow Y$ by means of a semantic link, whereas the evidence points towards a similar parallel process for $F \rightarrow G$ and $X \rightarrow Y$. We feel that this is an important distinction: whereas Holyoak & Thagard model the semantic similarity between SMALLER_THAN and LESS_MASS_THAN as a *sub-process* of analogy, it seems more likely that this similarity is computed by the *same process*, in parallel[2]. This can be seen most clearly via a fine-grained analysis of the predicates as in figure 3.

```
SMALLER_THAN
        RELATION(smaller_than)
        NO_OF_ARGUMENTS(smaller_than, 2)
        COMPARITIVE(smaller_than)
        NONREFLEXIVE(smaller_than)
        ASYMETRIC(smaller_than)
        TRANSITIVE(smaller_than)
        CONCERNS(smaller_than, size)

LESS_MASS_THAN
        RELATION(less_mass_than)
        NO_OF_ARGUMENTS(less_mass_than, 2)
        COMPARITIVE(less_mass_than)
        NONREFLEXIVE(less_mass_than)
        ASYMETRIC(less_mass_than)
        TRANSITIVE(less_mass_than)
        CONCERNS(less_mass_than, weight)
```

Figure 3: Predicate representations of SMALLER_THAN and LESS_MASS_THAN

Since from our perspective, this mapping process is what we mean when we talk about the 'analogical' process, it follows that there is no room for a semantic similarity constraint in a cognitive theory of analogy (or at least, it follows that before one can posit such a constraint, one will have to successfully distinguish 'analogy' from 'categorisation'). None of which means that we rule out the use of semantic links in mapping networks — rather, we suggests that they should be

[2]We are not saying that *all* similarities are calculated this way; what is true for relational predicates may not be true for others. It seems unlikely that structure mapping will feature in red vs. crimson.

seen as implementational details within a model, rather than embodiments of psychological theory.

One possible benefit to be gained from removing the semantic and pragmatic constraints[3] from Holyoak and Thagard's model is that it can allow a straightforward comparison between the structural elements in both theories to be made. Isomorphic mapping in multi-constraint theory differs from Gentner's structure mapping in a number of details; our next step is to evaluate the significance of these details by specifying both theories in a common, executable, specification language. Similarities and differences between the theories are currently blurred by the ambiguities of the language (English) in which they are stated. Whilst computational implementations of both theories exists, those implementations do not fully clarify either theory because they make no distinction between theoretical claims and implementational details. This is part of a broader issue: cognitive science is in need of tools and techniques in which to precisely state theoretical proposals so that their assumptions and implications are clear and comparison is possible (c.f. Cooper, Fox, Farringdon & Shallice, 1996). We are currently using the COGENT modelling environment (Cooper & Fox, 1997) to develop executable specifications of both structure-mapping theory and multi-constraint theory. It is our strong belief that this work will demonstrate that, modulo semantic similarity, the theories differ mainly in implemenation details.

The COGENT models will also provide a framework in which to explore more thoroughly issues of represention. We noted above the lack of focus on representation in analogical models (a point made frequently by Hofstadter, 1995). We have demonstrated that analogical process models can yield interesting results when applied to categorisation tasks (where perhaps too much focus is given to representation, c.f. Fodor & Lepore, 1996). Is it too much to hope that a similar cross-fertilisation might ultimately flesh out analogy's decidedly anorexic picture of representation?


## Acknowledgements

Many thanks to John Lee, Dedre Gentner and two anonymous reviewers for comments; thanks also to Richard White for interesting us in analogy in the first place.


## References


Clement, C.A. & Gentner, D. (1991) Systematicity as a selection constraint in analogical mapping, *Cognitive Science, 15*, 89–132.

Cooper, R. & Fox, J. (1997) COGENT: An interactive graphical environment for cognitive modelling. Unpublished: Manuscript available.

Cooper, R., Fox, J., Farringdon, J. & Shallice, T. (1996) A systematic methodology for cognitive modelling. *Artificial Intelligence, 85*, 3–44.

Falkenhainer, B., Forbus, K.D. & Gentner, D. (1989) The structure mapping engine: an algorithm and examples. *Artificial Intelligence, 41*, 1–63.

Fodor, J. & Lepore, E. (1996) The red herring and the pet fish: why concepts still can't be prototypes. *Cognition, 58*, 253–270

Forbus K.D., Ferguson, R.W. & Gentner, D. (1994) Incremental structure mapping. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, pp. 313–318.

Forbus K.D., Gentner, D. & Law, K. (1995) MAC/FAC: A model of similarity based retrieval *Cognitive Science, 19*, 141–205.

Gentner, D. (1983) Structure-mapping: a theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D. Ratterman, M.J. & Forbus, K. (1993) The roles of similarity in transfer. *Cognitive Psychology, 25*, 524–575.

Gentner, D. (1989) The mechanisms of analogical resoning. In S. Vosniadou and A. Ortony (eds.) *Similarity and Analogical Reasoning*. Cambridge University Press, Cambridge.

Hofstadter, D. (1995) *Fluid Concepts and Creative Analogies*. Basic Books, Harper Collins Publishers.

Holyoak, K.J. (1985) The pragmatics of analogical transfer. *Psychology of Learning and Motivation, 19*, 59–87.

Holyoak, K.J. & Thagard, P. (1989) Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295–355.

Holyoak, K.J. & Thagard, P. (1995) *Mental Leaps*. MIT Press, Cambridge, MA.

Keane, M., Ledgeway T. & Duff, S. (1994) Constraints on analogical mapping: a comparison of three models. *Cognitive Science, 18*, 387–438.

Medin, D. & Ortony, A. (1989) What is psychological essentialism? In S. Vosniadou & A. Ortony (ed.) *Similarity and Analogical Reasoning*. Cambridge University Press, Cambridge.

Nofosky, R.M. (1988) Similarity, frequency and category representations. *Journal of Experimental Psychology: Learning Memory and Cognition, 14*, 54–65.

Ramscar, M.J.A. & Pain, H. (1996) Can a real distinction be drawn between cognitive theories of analogy and categorisation? In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, pp. 346–351. San Diego, CA.

Ramscar, M.J.A., Lee, J.R. & Pain, H.G (1996) A cognitively based approach to computer integration for design systems. *Design Studies, 17*, 465–483.

Ramscar, M.J.A., Pain, H.G. & Lee, J.R. (1997) Some observed effects of systematic structure on categorisation decisions. Technical report, Department of Artificial Intelligence, University of Edinburgh.

Spellman, B. & Holyoak K. (1992) If Saddam is Hitler, then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62*, 913–933.


[3]The question of whether pragmatic constraints affect mapping or representation remains open — our comments regarding the relative similarity between SME and ACME when it comes to structural mappings might apply equally to pragmatic factors in ACME and I-SME.

# Examples And Generalisations: Using Surface Versus Structural Recall Biases To Probe Conceptual Storage

**Michael Ramscar, Helen Pain and Sarah Darrington (michael,helen,sarahd@dai.ed.ac.uk)**
Department of Artificial Intelligence, University of Edinburgh,
80 South Bridge, Edinburgh EH1, 1HN, Scotland
**John Lee (john@cogsci.ed.ac.uk)**
Human Communication Research Centre, University of Edinburgh,
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland

## Abstract

We argue that the key question in conceptual storage is best viewed not as a question of instances versus generalisations, but rather one of *unitary* versus *multiple* representation accounts of conceptual storage. On previous evidence, it has been difficult to determine whether a particular result stems from stored information regarding the concept or from the processes that operate in invoking a particular concept (Komatsu, 1992). In this paper, we attempt to shed some light on the nature of stored conceptual structure using the different influences that surface and structural features have been shown to have on the recall of a particular representation (Gentner, Ratterman and Forbus, 1993). We conclude that at least some concepts may not be stored using a unitary representation.

## Introduction

What is stored in the mind when a person learns a new category? What kind of mental representation (or representations) determine judgements of whether things are to be classed as bicycles, pianos, theories, or games, etc.? These questions lie at the heart of any understanding of human conceptual prowess, and a range of potential answers have been offered in their solution, ranging from necessary and sufficient essences, through defining prototypes, to averages of individual examples: the propositional schematic models of representation used to frame most theories of categorisation have the power to accommodate a combination of defining schemas and/or the examples from which schemas could be constructed with equal coherence, a flexibility that theorists have exploited to the full.

Whilst the work of Wittgenstein (1953) and Rosch, (1978) has tempered enthusiasm amongst researchers for classical essentialist accounts of conceptual storage, there has been much debate about the nature and plausibility of some form of prototype schema as a basis for conceptual storage (see Komatsu, 1992 for a discussion), though what is actually stored is the subject of much debate: Fodor and Lepore (1996) give many good reasons for doubting pure prototype theory, whereby just schemas are stored, whilst Nosofsky (1988) advocates a rejection of stored schemas altogether arguing that only examples are stored, and that new objects are classified by a process of comparison with stored examples. Other researchers favour some kind of theoretical framework as a basis for category storage (e.g. Medin and Ortony, 1989; Keil, 1989)

One point that has eluded many discussions of conceptual storage is that debates about the nature of representations revolve around questions of granularity rather then principle. As Borges (1962) succinctly demonstrates in the much quoted *Funes the Memorius*, even an 'instance' is a generalisation of sorts. When Funes struggles to

> 'comprehend that the generic symbol *dog* embraces so many unlike individuals... it bothered him that the that the dog seen at 3.14 (seen from the side) should have the same name as the dog at 3.15 (seen from the front).' Borges (1962) pp 93-94

he encounters the following problem: if the concept of *dog* determines whether things we encounter are dogs, then in the same fashion the concept *Spot* must similarly unite a certain class of experiences of a particular dog as being experiences of Spot the dog, and so on.

Once this factor is considered, we argue that the key question in conceptual storage is best viewed not as a question of instances versus generalisations, but rather one of *unitary* versus *multiple* representation accounts of conceptual storage. Unitary accounts of categorisation posit a single stored representation - schema, prototype or, perhaps, theory - in virtue of which items are classified into a category as the outcome of some process. Multiple representational accounts posit the storage of a number of representations, perhaps at different levels of granularity, from 'instances' (as Malt (1995) argues, regularities - perhaps as a result of perceptual constraints - are inevitable at some level) to broad intermediate generalisations, which may jointly or individually result in some object being categorised as the outcome of some process.

Different models of conceptual representation have different implications for theories of categorisation. If a unitary representation model is correct, then one would expect that provided one could specify the stored representation and the process by which objects were related to it, one should in principle be able to give a definitive account of, say why it is some things are X's. On the other hand, a multi-representational account might not admit any definitive account of X's at all, since the multiplicity of relations between the differing stored representations that could influence an object's X-ness might preclude any kind of general account. A specific object's X-ness might be dependent upon that - and only that - objects' interaction with a particular subset of the stored elements relating to X-ness, and the concomitant process by which X-ness is adjudged.

A problem in much research into categorisation is that experimental results have rarely - if ever - directly indicated anything about conceptual representation. As Komatsu (1992) concisely notes, it is difficult to determine whether particular results stem from stored information regarding concepts (e.g. propositional or imaginal information) or from the processes that operate in invoking a particular concept (c.f. Smith and Medin, 1981). The vast majority of theories discussed in Komatsu's (1992) comprehensive review assume a 'straight' unitary representation, the exception being granular instance-based approaches to categorisation, and perhaps explanation-based approaches (although the lack of any specific formulation of an explanation (or theory) based model of categorisation makes it impossible to deduce the kind of stored representations such a theory would entail).

In the light of these considerations, we feel that direct evidence regarding the nature of the storage of a concept, or concepts, could have important implications for the way categories are viewed: especially if that evidence fails to support a unitary-representation account.

## A process model

In previous work (Ramscar and Pain, 1996) we have questioned an intuitive distinction - held by most psychological researchers - between analogy and metaphor on one hand, and categorisation on the other. We have argued that although one might ordinarily distinguish between category membership and analogy according to realist terms, there are good reasons for treating this distinction with caution at a theoretical level, especially when we focus upon cognitive processes.

The argument for this is twofold. Firstly, in principle: any intuitive, pre-theoretical distinction between real (intra-categorical) and metaphorical or analogical (inter-categorical) cognisance hinges upon a contrast definition that sits ill with what we know empirically and theoretically about categorisation. Analogy and metaphor are consistently defined in contrast to categorisation (e.g. Holyoak and Thagard, 1995); yet a contrast definition relies on an account of the element with which the contrast is to be drawn: the message of many years of empirical and theoretical work (Wittgenstein, 1953; Rosch, 1978; Barsalou, 1983; Murphy and Medin, 1985) is that there is no clear account of intra-categorical associations between individual items with which supposedly inter-categorical associations can be contrasted. Our dubiety towards this distinction is further borne out by analyses of reaction time studies. For instance, Hoffman and Kemper (1987) review of a number of reaction time studies amply demonstrates the absence of evidence for (and amount of evidence against) the widely held belief that literal (intra-categorical) meanings are processed faster than metaphorical (inter-categorical) meanings (Récanati (1995) also reviews some interesting evidence contra the 'two-process' approach).

Further support for this position comes from an earlier study (Ramscar and Pain, 1996) which used Gentner's (1983) Structure Mapping Theory (SMT) to address the question of whether analogy can be distinguished from categorisation by contrasting categorisational with analogical processes.[1] Gentner's theory proposes that in mapping and inference between two representations, systems of relations will be mapped in preference to individual features, and that higher order relations - relations between relations - will be mapped in preference to lower order and first-order (relations between objects) relations: the *systematicity principle* (Gentner, 1983).

Ramscar and Pain's (1996) subjects were presented with Gentner et al's materials and asked to categorise them. It was expected that structure mapping would determine categorisation as well as analogy: i.e. Gentner's theory assumed that match items with only structural similarities ('analogues') belong to different categories: Ramscar and Pain predicted that they would be 'categorised' together. Ramscar and Pain found that a significant majority of the groupings formed by subjects had only shared systematic structure (traditionally defined as analogy) as a common feature amongst members of the categories formed, and that the process underlying subjects' categorising could be modelled - and hence predicted - by SMT; they concluded that there was nothing to distinguish subject's analogical from their categorisational behaviour. These results were replicated by Darrington, Lingstadt and Ramscar (1998) who discovered that subjects would even use shared structure in novel object descriptions to *over-ride* existing category groupings.

Our second argument is a pragmatic one. Analogy models can provide a framework through which the interactions between representation and process can be predicted and observed. Traditionally categorisation models have concentrated on object descriptions, making use of very representationally-simple attribute-value lists (Murphy & Medin, 1985), whereas analogy research has examined relationships between highly structured representations (considering the influence of attributes, relations and higher-order relations in judgements of similarity). Research into analogy differs from research into categorisation is in the richness of its process models. A number of detailed, plausible, well accepted models of the analogical process exist (Holyoak and Thagard, 1995; Forbus, Gentner & Law, 1995): the same cannot be said of existing categorisation models. Research in analogy has focused upon the processes of mapping and inference between representations, and the interplay between this process and those representations rather than simply determining what kind of mappings certain static representations allow, a characteristic of much categorisation research (c.f. Komatsu, 1992).

## Representation and process

A series of studies by Gentner, Ratterman and Forbus (1993) has explored the factors determining how items are accessed, i.e. how representations are selected in order to allow similarity mapping to take place. These have shown that access relies primarily upon surface attribute (or object attribute) matches, and they propose that the

[1] Though developed as a theory of analogy, SMT is now considered by Gentner to apply more generally as a theory of similarity.

process underlying judgements of similarity can be decomposed into two sub-processes:

- *Accessing* a similar (*base*) situation from memory, based primarily on surface similarity
- Creating a *mapping* from base to target using structural commonalities (SMT).

We believe that this process model can offer a solution to the difficulties, mention above, of determining whether particular effects result from stored representations or from the processes that operate in invoking representations. The following experiment was designed to use this detailed model of the processes that determine the retrieval and mapping (and classifying) of representations to empirically probe the nature of conceptual storage.

## The Experiment

In the course of Ramscar and Pain's (1996)[2] classification experiment, subjects were asked to give each of their classes a name that was meaningful to them. Because of the particular nature of that task, this involved subjects developing (and learning) 'categories' that contained some items that had only structural relations in common. By examining the attributes they could recall that were associated with that name, we aimed to use Gentner et al's findings about systematic structures vs. attributes to determine the representation associated with the name. If subjects stored some kind of abstracted prototype - i.e. a unitary representation of their category - we would expect that the attributes associated with the most prototypical stories would be most readily retrieved from memory, with attributes recalled insofar as they were relevant to the prototype (perhaps along the frequency lines one might expect from the analysis in table 1).

*Story 1 - Base story*

Once there was a teacher named Mrs. Jackson who wanted a salary increase. One day, the principal said that he was increasing his own salary by 20 percent. However, he said there was not enough money to give the teachers a salary increase.

When Mrs. Jackson heard this she became so angry that she decided to take revenge. The next day, Mrs. Jackson used gasoline to set fire to the principal's office.

**Then she went to a bar and got drunk.**

*Story 2 - Literal similarity*

Professor Rosie McGhee very much wanted a raise. One day the provost announced that he was giving himself a raise. However, he said that since money was short, no one else would get a raise this year.

After Professor McGhee heard this she became so upset that she decided to get even. One hour later, Professor McGhee blew up the administration building with dynamite.

*Story 3 - True Analogy*

McGhee was a sailor who wanted a few days of vacation on land. One day, the captain announced that he would be taking a vacation in the mountains. However, he said everyone else would have to remain on the ship.

After McGhee heard this he became so upset that he decided to get revenge. Within an hour McGhee blew up the captain's cabin with dynamite.

**Figure 1:** Sample stories from Gentner, et al (1993) - the text in bold

We did not expect this to happen There is too much empirical and theoretical work that cannot be accommodated by a unitary representation account (Wittgenstein, 1953; Rosch, 1978). Our hypothesis was that subjects would not have abstracted a unitary representation from the stimuli that they had classed together, but would instead have stored a number of representations that they associated with the relevant class name. We predicted that introducing a class name out of context would make it equally likely that any given stored representation associated with that name would be recalled, with the representation recalled initially driving further recall. Since our model of recall is feature driven, we expected results to polarise, with those representations with few surface attributes in common leading to minimal recall of oneanother, and those with many surface attributes in common leading to good recall of oneanother

### Subjects

The subjects were 20 volunteers, a mixture of postgraduate and undergraduate students from the Artificial Intelligence Department at the University of Edinburgh.

### Materials

Materials were the classified sets of "Karla the hawk" stories (Gentner, Ratterman & Forbus, 1993; see figure 1 for examples) produced by the subjects in Ramscar and Pain (1996). Each set contains 3 stories, and taking one story as a base, the following taxonomy of similarity relationships between the stories in a set can be defined:

- *Literal similarity* (LS) matches to base include both common relational structure and common object descriptions;
- *Structural similarity*, (SS) match based upon a common system of internal relations with base.

Thus each set comprised a base (B), a story literally similar to it (LS), one structurally similar to it (SS - i.e. with no object attributes in common with the base, although SS stories did share some object attributes with LS). In order to determine the relative effects of object versus structural matches in this experiment, the "Karla the Hawk" stories were analysed and rated to determine the level of attribute commonalities between the individual stories in each set.

Two raters gave a numeric value to each of a range of possible surface attribute correspondences between stories (see figure 2 for details), and then individual attribute correspondences were totalled and averaged between the two raters in order to determine the overall correspondences between stories (table 1). Consistency between raters was 82.5%. Differences between raters were resolved by discussion.

| Correspondence | Example | Value |
|---|---|---|
| 1 -1 map | man, man | 5 pts |
| strong map | street, road | 4 pts |
| 'analogical' association | conned, robbed<br>fireman, paramedic | 3 pts |
| weak 'analogy' | enlisted, begged<br>fireman, nurse<br>disappeared, shattered | 2 pts |
| weak association | Fred, Bill (both names)<br>fireman, accountant (both jobs) | 1 pt |

**Figure 2:** Classification criteria for determining surface similarity.

[2] The experiment discussed here was conducted alongside Ramscar and Pain's study; this is the first time it has been analysed and reported.

| Set | B-LS | LS-SS | SS-B |
|---|---|---|---|
| 1 | 100 | 90 | 85 |
| 2 | 100 | 100 | 85 |
| 3 | 100 | 65 | 45 |
| 4 | 100 | 55 | 45 |
| 5 | 100 | 85 | 55 |
| 6 | 100 | 85 | 75 |
| 7 | 100 | 75 | 60 |
| 8 | 100 | 70 | 45 |
| 9 | 100 | 85 | 65 |
| 10 | 100 | 75 | 60 |
| 11 | 100 | 95 | 75 |
| 12 | 100 | 75 | 60 |
| 13 | 100 | 75 | 50 |
| 14 | 100 | 95 | 70 |
| 15 | 100 | 50 | 35 |
| 16 | 100 | 95 | 75 |
| 17 | 95 | 100 | 70 |
| 18 | 100 | 80 | 55 |
| 19 | 100 | 80 | 60 |
| 20 | 100 | 100 | 60 |

**Table 1:** Object attribute (surface feature) similarity ratios between stories by set.

## Procedure

During the classification task (Ramscar & Pain, 1996, see above), subjects were asked to give each of the classes they produced a name that 'would be meaningful to them later'. After finishing the classification task, each subject was given a 5 minute break, and then undertook a 20 minute diversionary task (searching for post-codes from a directory) before being given another 5 minute break. The same subjects were then presented with a sub-set (usually 4) of the names they had assigned to classes during the classification task, and asked to 'write down what you can remember about the various features (or you may like to see them as attributes) of each of the scenarios associated with each name. E.g. you may have had a scenario about a door that needed varnishing. Features, or attributes, associated with such a scenario would be "door" and "varnishing"'. Subjects were given 10 minutes to complete the task.

## Results

The 20 subjects yielded a total of 70 recall episodes.

*Scoring*

The recalled features were evaluated by two judges using the same scale that was used to evaluate feature correspondences between stories (see figure 2). Points were awarded for correspondences between features named by the subjects and the actual feature name in each given story in order to get a reflection of the accuracy of each individual recall episode. The total attribute recall for each story was calculated and averaged between raters. Consistency between raters was 84.1%; as in the rating of story commonalities, differences between the raters were resolved by discussion.

*Individual Story Recall*

B was best recalled story for 37% of all recall episodes, SS in 30% of cases and LS in 33%; as predicted, there was no significant bias towards recalling any particular type of story. However, when we looked at the pattern of recollection, irrespective of the particular stories each subject had recalled, there were significances in the quality of recall between the best recalled and the next best, and the next-best and the worst recalled stories. Subjects tended to clearly recall one story better (70 cases, Mean ($M$) = 19.75) than the next ($M$ = 14.74), (within groups $t(69)$ = 8.846, $p$ < 0.0001), and then these next best recalled stories better than the worst ($M$ = 11.1), $t(69)$ = 11.802 $p$ < 0.0001.

*Individual Recall Orderings By Story Type*

<u>Base (B)</u>

In cases where B was the most recalled story in terms of features (26 cases, $M$ = 18.44), there was a significant difference in the quality of recall over the next best recalled story type, LS (M = 14.83), $t(25)$ = 4.434 p < 0.0001, which in turn was recalled significantly more than SS (M = 10.15), $t(25)$ = 3.77 p < 0.0001.

<u>Literal Similarity (LS)</u>

When LS was the best recalled story (M = 20.78), next best was B (23 cases, $M$ = 15.85), $t(22)$ = 5.288 p < 0.0001, with SS recalled least well ($M$ = 11.96), $t(22)$ = 2.095 p < 0.05.

<u>Structural Similarity (SS)</u>

In cases where SS stories were qualitatively best recalled (21 cases, $M$ = 20.12), the next best recalled type was LS, ($M$ = 13.43), $t(20)$ = 5.886 p = < 0.0001, and the least recalled type was B ($M$ = 11.36), although the between type difference between LS and B was not significant ($t(20)$ = 1.901 p < 0.072); the difference in recall quality between SS and B was still significant ($t(20)$ = 6.609 p < 0.0001).

| B Best Recalled | Next Best - LS | Next Best - SS |
|---|---|---|
| 26 cases; Mean = 18.44 | Mean = 14.83 | Mean = 10.15 |

| LS Best Recalled | Next Best - B | Next Best - SS |
|---|---|---|
| 23 cases; Mean = 20.78 | Mean = 15.85 | Mean = 11.96 |

| SS Best Recalled | Next Best - LS | Next Best - B |
|---|---|---|
| 21 cases; Mean = 20.12 | Mean = 13.43 | Mean = 11.36 |

**Table 2:** Mean recall orderings by story type

## Discussion

The experiment produced little evidence to support a belief that our subjects had abstracted and stored schemas from the groups they had classified, despite the fact that a shared structural schema was the basis of subjects' original classification decisions (Ramscar & Pain, 1996). If some version of a stored prototype theory were true, we expected a majority of LS features to be recalled in most instances. In fact, B features were most often recalled, though not significantly: the trend favoured a random distribution. Another result that might also favour prototype theory would have been a situation where all the stories were recalled with much the same frequency, i.e. LS=B=SS, since such a result could be a product of the strong feature commonalities between the LS stories and members of both of the other story types. However, there was a significant trend for subjects to recall one story more than another, and the next best story more than the least recalled story.

One criticism of this initial conclusion might be that our task, by asking subjects to write down features associated with individual scenarios, biased against recall of a single unitary or abstracted representation. However, again, given that LS story set attributes formed the intersection of the sets of features for *all* of the story sets, we would expect any unitary or abstracted representation of all of the stories in a given set to contain mainly LS story set attributes and to cause a majority of LS story set attributes to be recalled. Clearly, this was not the case.

If subjects randomly recalled an individual instance of a class then we would expect from Gentner et al's similarity recall findings that the attributes of this instance should influence which other story they might recall from the class: if B is recalled, recall of a B story should prompt recall of an LS story over a SS story, as B shares more surface attributes with LS than SS (B and LS share 10 surface attributes to every 6 shared by B and SS), and Gentner et al's findings were that surface attributes rather than shared structure promote recall: where B features were best recalled this pattern emerged throughout our study. As we predicted, as a result of LS stories sharing a higher percentage of surface attributes with B than SS shared with B, recall of B led to a significantly higher quality of recall for LS than SS. The results where LS stories were most strongly recalled also supported this analysis, with SS recall prompting significantly better recall of B attributes than SS attributes (even though, comparatively, LS shared more surface features with SS than SS-B).

Indicative of the fact that SS shared fewer surface commonalities with the other story types, (SS-LS attribute commonalities were much weaker than B-LS, and in specific sets little greater than B-SS; see table 1), good SS recall did not produce a bias towards LS or B as the next-best recalled story type; although our results showed some tendency towards SS prompting LS over B, it was not significant (see also table 2).

Our hypothesis that subjects would recall stories individually from memory is further supported by the nature of subject's recollection. Irrespective of the particular stories each subject had recalled, there were significances in the quality of recall between the best recalled, and the next best, and the next-best and the worst recalled stories.

A possibility that we cannot eliminate in this instance is that subjects may have been influenced by the order in which stories were presented. In our study, presentation was randomised in a manner that made it impossible to correlate presentation and recall orderings. However, as we noted above, there is a very strong correlation between the features of the LS story in each set and the features of an idealised prototype of that set (i.e. LS features, especially at the level of names of relations, are most typical of all members). Since subjects, who used shared structure to group the stories, did not show a tendency to recall LS features best, any correlation between presentation order and recall order could only further support the case against some shared representation having been abstracted and stored.

Applying Gentner, et al's (1993) persuasive analysis of the influence of surface vs. deep structure on recall to our results, it would appear that being presented with a class name in no particular context caused the subjects to randomly recall one of the examples associated with that name, and then use that example as the stimulus for recalling other class members. On this evidence, it would appear that subjects had stored class examples along with a cue - the class name - rather than any generalisation of the class itself. These findings suggest that at least some concepts are stored as multiple-representations, rather than as unitary conceptual schemas.

## General Discussion

We have shown in this paper that subjects if subjects name a class whose membership is determined entirely by shared structure, they can retrieve information regarding its members without appearing to abstract a common schema definitive of that class. Such evidence for a non-unitary account of category representation should not come as a surprise: the literature is filled with material that casts doubt of the plausibility of unitary accounts of concept representation (Wittgenstein, 1953; Rosch, 1978; Komatsu, 1992).

Our interpretation of these findings does not however lead us automatically to the view that 'category' associations are exclusively driven by examples; rather, we subscribe to the view put forward by Wittgenstein (1953; see also Ramscar, 1997). Wittgenstein provides a number of good reasons to believe that human 'categories' cannot be given simple unitary accounts that are amenable to definition by a single schema; he also questions what could be intrinsic to a 'generalised' category schema that would cause it to be used differently to an example of that which it was supposed to be a generalisation of. We find these questions compelling. At the same time, experience, intuition and a respect for natural parsimony makes us wary - though in practice rather than in principle - of jumping to the conclusion that all categorical associations are a sum of processed similarities with all stored exemplars. We would argue, however, that progress is not contingent on finding a solution to the question of exactly how categories are represented. It is the *process* of association that is important to us, and from a processing point of view, the question of whether a stored schema relates to an example or some intermediate generalisation is not necessarily important. What *is* important, from the point of view of processing, is to show how stored representations (whatever their exact nature) relating to the individual surface elements (the feature 'nodes' in SMT) comprising the schemas that represent items being associated in the manner described in structure mapping theory are recursively processed in turn, such that the 'network of similarities that determines a given association can settle (see Ramscar, Pain and Cooper (1997) for an illustration of our recursive view of the associative mapping process).

Another advantage of viewing categorisation from a process oriented perspective is that it offers up the possibility of bridging the divide between two seemingly conflicting accounts of categorisation: similarity-based accounts, and more rationalist 'theory-based' accounts. Whilst similarity-based accounts of categorisation (e.g. Rosch, 1978) can capture much of the nature of our long-term categories (that they have no explicit definitions, and that there are usually a number of properties that are generally associated with 'categories'), they have been criticised for failing to capture the explanatory power of categories, and for failing to explain why people may categorise in ways that go beyond surface similarity (Medin and Ortony, 1989). Keil (1989) notes that whilst children's concepts are based upon pure similarity - 'original sim' - these then get replaced with more theoretically based conceptual understandings as a child develops (see also Murphy and Medin, 1985; Medin and Ortony, 1989).

What these theory based accounts lack is any clear description of the process by which theories are supposed to govern categorisation, or, importantly, exactly what kind of theoretical understandings are supposed to underpin human conceptual understandings which are notoriously vague and often incoherent. We believe that Gentner's (1983) insight, that relations between features (i.e. *constraints*) can drive computations of similarity, can go at least some way towards squaring this apparent circle. The assumption that similarity cannot account for the apparent theoretical notion of categorical associations stems from the assumption that similarity is driven by features alone. When the relational constraints that drive the similarity comparison process are considered, and once a richer notion of mental representation than unrelated feature clusters is included in the picture, it is possible to see how similarity *can* provide an account of the associative process that determines human categorisation decisions which capture their theoretical flavour,[3] *without* entailing theoretical accounts of individual 'categories' which by their nature (c.f. Wittgenstein, 1953) are more of the nature of underdetermined collections of generalisation, similarity and history than they are unitary 'theoretical' entities.

## Acknowledgements

The authors would like to thank: Richard Cooper and two anonymous reviewers for their comments; and Dedre Gentner for her assistance in providing our materials.

## References

Barsalou, LW (1983) Ad hoc categories *Memory and Cognition*, **11**:211-227

Borges, JL (1962) *Labyrinths* Penguin, New York, NY

Darrington, S, Lingstadt, T and Ramscar, MJA (1998) Analogy as a sub-process of categorisation. *Submitted.*

Fodor, J & Lepore, E (1996) The red herring and the pet fish: why concepts still can't be prototypes. *Cognition* **58**:253-270

Forbus K, Gentner, D, & Law, K (1995) MAC/FAC: A model of similarity based retrieval *Cognitive Science* **19**:141-205

French, RM (1995) *The Subtlety of Sameness*, MIT Press, Cambridge, Ma.

Gentner, D (1983) Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, **7**: 155-170.

Gentner, D, Ratterman, MJ & Forbus, K (1993) The roles of similarity in transfer. *Cognitive Psychology* **25**: 524-575

Hoffman, R R and Kemper, S (1987) What could reaction-time studies be telling us about metaphor comprehension? *Metaphor and Symbolic Activity* **2(3)**, 149 - 186

Holyoak, KJ and Thagard, P (1995) *Mental Leaps*. MIT Press, Cambridge, Ma.

Keil, FC (1989) *Concepts, kinds and cognitive development.* MIT Press, Cambridge, Mass.

Komatsu, LK (1992) Recent views of conceptual structure. *Psychological Bulletin*, **112(3)**, 500-526

Lakoff, G and Johnson, M (1980) *Metaphors we live by*. University of Chicago Press, Chicago, Ill.

Malt, BC & Smith, EE (1984) Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behaviour* **23**:250-269

Malt, BC (1995) Category coherence in cross-cultural perspective. *Cognitive Psychology* **29**:85-148

Murphy, GL and Medin, DL (1985) The role of theories in conceptual coherence. *Psychological Review*, **92**, 289 - 316

Nofosky, R M. (1988) Similarity, frequency and category representations. *Journal of Experimental Psychology: Learning Memory and Cognition* **14**:54-65

Ramscar, MJA & Pain, HG (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? *(In Proc.) 18th Annual Conference of the Cognitive Science Society.* LEA, New Jersey.

Ramscar, MJA, Pain, HG & Cooper, R (1997) Is there a place for semantic similarity in the analogical mapping process? *(In Proc.) 19th Annual Conference of the Cognitive Science Society.* LEA, New Jersey.

Ramscar, MJA (1997) Wittgenstein and the nature of psychological categories *Proceedings of SimCat 97*, Dept. of Artificial Intelligence Conference Proceedings, University of Edinburgh, Scotland, 205-211.

Récanati, F (1995) The alleged priority of literal interpretation *Cognitive Science* **19**:207-273

Rosch, E, Simpson, C. & Miller, RS (1976) Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance* **2:4**:491-562.

Rosch, E (1978) *Cognition and Categorisation.* Lawrence Earlbaum Assoc., New Jersey.

Smith, EE, Shoben, EJ, & Rips, LJ (1974) Structure and process in semantic memory: a feature model for semantic decisions *Psychological Review* **81**:214-241

Smith EE and Medin D (1981) *Categories and concepts.* Harvard University Press. Cambridge, Mass.

Wittgenstein, L *translated by* Anscombe, E (1953). *Philosophical Investigations* Blackwell, Oxford.

---

[3] Keil (1989) rejects SMT as a basis for such an explanation because it gives no privileged status to the *causal* constraints that he argues are most important to category judgements. We find this mystifying, since SMT maps according to the most weighted constraints embodied in representations of items in working memory: if causal constraints are privileged, one would expect this to be reflected in the weighting of mental representations. It is a strength of SMT that by *not* favouring causal constraints it does not preclude categorical associations formed on the basis of other constraint systems.

# Analogy As A Sub-Process Of Categorisation

**Sarah Darrington, Toby Lingstadt and Michael Ramscar**[†]
**(sarahd, tobyl, michael@aisb.ed.ac.uk)**
Department of Artificial Intelligence, University of Edinburgh,
80 South Bridge, Edinburgh EH1 1HN, Scotland

## Abstract

Analogy has traditionally been defined in terms of a contrast definition: analogies represent connections between things which are distinct from the 'normal' connections determined by our 'ordinary' concepts and categories. In this paper we present empirical evidence which, when added to other findings, supports our argument that in the light of current knowledge, the distinction between the two is based more on folk-psychology than on empirically based theory.

Research into analogy is however, distinct from research into categorisation when it comes to the richness of its process models. A number of detailed, plausible models of the analogical process exist (Forbus, Gentner and Law, 1995; Holyoak and Thagard, 1995): the same cannot be said of categorisation. On the other hand, these analogical process models make a number of explicit and implicit assumptions regarding an 'external' categorical process. Whilst treating these processes as separate has been useful in constraining the scope of cognitive investigations, we argue that it ultimately confuses the relationship between analogy and categorisation and is hampering the progress towards further understanding of both.

## Introduction

The belief that analogy and categorisation are distinct and separable cognitive processes has widespread appeal. In our ordinary lives we unquestioningly accept an ontology that distinguishes between literality - saying what something 'really' is - and analogies and metaphors, which, however informative they may be, are nevertheless not considered to be real statements about the world. We might talk of "the *foundations* of a theory"; we might wish to "*buttress* a theory with more facts"; "theories that we *construct* can also *collapse*", but from our everyday viewpoints, an igloo and a castle and a skyscraper *really* are similar in a way that similarities between buildings and theories are not; we can talk of someone's foxy cunning without *really* meaning to directly equate the cognition of fox and human cunning. French (1995) describes an experience of suggesting to an academic audience that an upturned orange-crate, when covered with a cloth and laid out with a picnic, might *really* be described as a table. This met with the swift response, "An orange crate is an orange crate is an orange crate." The attachment to pre-theoretical intuitions is a strong one, even amongst those who seek to explore them.

Research into categorisation, analogy and metaphor has accepted this realism, as indeed for a large part has cognitive science in general. Holyoak and Thagard (1995) describe a world in which "we think we see things as they really are", and analogy is used in order to recycle our existing knowledge of the real world to formulate new bits of 'real' knowledge.

Ramscar and Pain (1996) questioned the basis of these everyday distinctions in theory by querying the formulations that are offered in their defence. Analogy and metaphor are defined in contrast to 'categories' - Holyoak and Thagard (1995) describe analogy and metaphor as something that "connects two domains in a way that goes beyond our normal category structure" (pp 217) - a definition that does nothing definite by way of defining when we consider the account that can be provided of what constitutes a 'normal category structure' (c.f. Komatsu, 1992).

Empirical evidence was introduced to support this dispute regarding current definitions: Ramscar and Pain examined whether analogy could be distinguished from categorisation by contrasting the two processes. They presented participants with Gentner, Ratterman and Forbus' (1993) classic analogy materials (the 'Karla the Hawk' stories) and asked them to categorise them. Given that Gentner *et al* define the analogical mechanism in terms of structure mapping theory (see below for a full account), Ramscar and Pain hypothesised that assuming that the 'analogical' process was not distinct from the basic categorisation process, structure mapping would determine categorisation.

Gentner *et al* assumed that match items with only structural similarities (i.e. analogues) should be considered as belonging to different categories. Ramscar and Pain predicted that they would be categorised together. They found that 79.5% of the groupings formed by participants in their study had only shared systematic structure (traditionally defined as analogy) as a common feature amongst members of the categories formed. In contrast, only 5% of groupings produced had common object descriptions as the common similarity across categories (i.e. the attribute matches often thought to be determinate of categorisation). To the 79.5% of structural congruity groupings could be added a further 8% of classifications where structural *additions* to otherwise structurally congruent representations caused them to be classed singularly. Ramscar and Pain concluded that that mechanisms normally considered to be analogical - specifically the preference for mapping systematically similar structures - could also in fact support categorisation

[†] This ordering is alphabetical: the authors all contributed fully to the experiment and the composition of this paper.

tasks, and that in their study no discernible difference could be found between analogical and categorical behaviour.

Ramscar and Pain's study can be added to other theoretical and empirical evidence which casts doubt on a two-process account of literal (categorical) versus non-literal (analogical or metaphorical) reasoning, such as Hoffman and Kemper's (1987) review of a number of reaction time studies which also convincingly demonstrates the paucity of the evidence for the widely held belief that literal (intra-categorical) meanings are processed faster than metaphorical (inter-categorical) meanings (as well as the considerable evidence for the opposite effect; see also Récanati, 1995).

Despite the weight of evidence contra the two-process account - and the concomitant lack of evidence for it - suggestions which violate the two processes-account still tend to meet with some incredulity (c.f. French, 1995, above). The belief that an orange crate is an orange crate is an orange crate holds great sway. Indeed, such is the two-process account's entrenchment in ordinary, pre-theoretical understandings of the world that a more than usual quantity of counter-evidence seems to be required simply for the two-process account to be subjected to the usual standards of scientific and theoretical justification. And, whilst Ramscar and Pain's (1996) study might apply to classifying *stories* their study provides little evidence that this classification of stories can be generalised to other categorisation behaviour.

## Structural systematicity and 'normal' categories

The following experiment was designed to offer more evidence for Ramscar and Pain's (1996) claim that a sub-process rather than two-process view should be taken of the relationship between analogy and categorisation (and thus add further support to Medin, Goldstone and Gentner's (1993) contention that structure-mapping may play an important part in categorisation). The Ramscar and Pain study utilised the 'Karla the Hawk' story sets (Gentner at al, 1993). These were designed as materials for studying analogy, and comprised a number of scenarios, typically episodic stories, with a controlled variance of relations between the narrative features of each story set member. In order to demonstrate the generalisability of Ramscar and Pain's findings, we designed a new set of materials, based on the 'Karla the Hawk' sets, but which rather than being stories, were sets of descriptions of novel objects (again, with a controlled variance of relations between the narrative features of each set member; see Figure 1).

The principle reasoning behind this is straightforward: typical 'natural' categories - the kind of categories found and used in human societies - tend to concern objects - and other, more regular 'things in the world' - rather than stories (Rosch, 1978). Classifying objects is more akin to ordinary categorisation than classifying stories (though a set of descriptions of 'rituals' was also included to reflect the fact ordinary categories reflect a good deal more than just physical regularities). Thus, the resulting object descriptions fell neatly and clearly into 'normal' categories (Figure 2).

Ramscar and Pain showed that shared structural systematicity (Gentner, 1983; see below) - typically defined as analogy - was the key determinant in participants' categorising in their study. In the light of this finding, we hypothesised that because structural commonalities in the object description sets ran across the 'normal' categories embodied in the set's object descriptions, these 'normal' category boundaries would be ignored as participants categorised objects according to shared structure.

**Base**

A Karla is a novel type of cooking pot, used by the Timuni in Alnata.

The structure of the Karla is designed in order to reduce the heat inside, and therefore prevents the food getting burned in the scorching cooking fires.

Water is poured into a layer of the Karla during cooking, which cools the food.

**Literal Similarity**

The Valkri is a special kind of frying pan, used by the Jalpeni in Frodon.

The Valkri is created in such a way as to be able to reduce heat, thereby preventing meat being getting burned when using the extreme temperatures of the cooking fires.

A liquid is poured into the layers of the frying pan when cooking, which cools the temperature of the meat.

**Structural Similarity Only**

The Vubu is a special wall built by the Jakar tribesmen in Frodon.

The Vubu is built in such a way as to be able to reduce the heat within it, thereby preventing the Jakar from sweating too much in the extreme temperatures of the midday sun.

A liquid is pumped through the Vubu, which cools the stone and therefore prevents the Jakar within the walls from getting too hot.

**Mere Appearance**

The people of Frodon use a special type of frying pan, known as the Valkri.

The Valkri is designed in order to allow it to be handled by children, as this can be difficult.

Its handle is designed with a special U-shape, which enables it to be held by people with small hands.

**Structurally Similar to MA**

The Jakar tribesmen of Frodon have built a special wall known as the Vubu.

The Vubu's stone gates can be opened by elderly people, despite their heavy weight.

Handles set in the wall incorporate springs, which allow weaker people to open the gates.

**Objects Only**

A new type of cooking pot, called the Karla, is used by the people of Alnata. Karlas can be purchased in a range of colours. Food cooked in a Karla tastes great.

**Figure 1** Example of an object description set used in the study.

| Set Member Description | | | | | | |
|---|---|---|---|---|---|---|
| Object Sets | B | LS | SSO | MA | SMAO | OO |
| SET 1 | plant | plant | tribe | plant | tribe | plant |
| SET 2 | country ruler | country ruler | leading animal | country ruler | leading animal | country ruler |
| SET 3 | board game | board game | field game | board game | field game | board game |
| SET 4 | animal | animal | tribe | animal | tribe | animal |
| SET 5 | cooking utensil | cooking utensil | wall | cooking utensil | wall | cooking utensil |
| SET 6 | animal | animal | priest | animal | priest | animal |
| SET 7 | chant | chant | game | chant | game | chant |
| SET 8 | food | food | drink | food | drink | food |

**Figure 2:** 'Normal' categories embodied in each object description set.

## Gentner's structure mapping theory of similarity

Gentner's (1983; Falkenhainer, Forbus and Gentner, 1989) Structure Mapping Theory (SMT) is well known, and we include only a brief account of the most salient aspects here.

SMT proposes that the mapping and inference between two representations can be achieved by assigning correspondences between objects and attributes and then mapping predicates with identical names. In order to do this, Gentner assumes a predicate-like representation distinguishing between *objects, object-attributes* and *relations*. Object-attributes are those predicates that have one argument and describe object properties. Relations are divided into a hierarchy of orders, with those predicates with two or more arguments which are used to describe relations between objects forming the lowest order, and those predicates describing different levels of relationships between relations forming the higher orders.

The theory itself comprises two parts: *mapping rules* and the *systematicity principle.* Mapping rules state that (a) attributes of objects are not mapped and (b) relations between objects are preserved. The systematicity principle requires that higher order relations are mapped preferentially, followed by the relations that constitute the higher order arguments. Ramscar and Pain (1996) showed that participants' classification of stories could be predicted and explained according to SMT.

## The Experiment

### Participants

20 volunteers participated in this experiment. The participants were a mixture of Artificial Intelligence and Psychology students from the University of Edinburgh.

### Materials

The basic materials used for this study were 8 sets of 'Karla the Pot' novel object descriptions (see Figure 1 for examples). These were descriptions of objects created to replicate the framework used by Gentner, Ratterman and Forbus (1993) in the creation of the "Karla the Hawk" stories.

As with the materials used by Gentner et al. the following taxonomy of similarity relationships between the object descriptions was defined:

- *"Literal similarity"* matches include both common relational structure and common object attributes;
- *"Surface matches"* are based upon common object attributes, plus some first order relations;
- *"Structural similarity"* matches are based upon a common system of internal relations;
- *"First order"* matches only have first order relations as a common feature;
- *"Object only"* matches only have object matches in common between the object descriptions.

Each of our sets consists of a base (B), a literally similar object description (LS), an object description that shared the same structure as the base, but no object attributes (SSO), a mere-appearance object description, with surface and first order commonalities with the base (MA), an object description which shared structure with the MA, and object attributes with the SSO (SMAO), and an object only match object description, with only surface attribute commonalities with the base (OO). This allowed for a number of potential groupings to be formed, according to the classification strategy participants adopted.

We predicted that despite the fact that we were using novel object descriptions which embodied existing categories rather than Gentner et al's relatively 'category-neutral' stories, participants would again use structural similarity as their categorical similarity determinant, putting analogues and bases into the same categories (i.e. B, LS and SSO together), rather than grouping match items at the object level (i.e. grouping B, LS, MA and OO together; which also equated to existing category membership; see figure 2).

### Procedure

Each participant was presented with eight envelopes, each containing a different set of six novel object descriptions, and was asked to work through them one set at a time. Sets were presented in random order, as were the object descriptions within them.

Participants were instructed to read through the object descriptions within a set several times, until they felt familiar with their contents. They were then asked to put the objects together into groups, grouping the things that fitted most naturally together in their judgement. Groupings could range from putting all descriptions into the same group to having them all in separate groups as well as all variations in between.

When the categorisation decisions had been decided on, the participant pasted them onto a large sheet of blank paper and then circled each grouping using a marker pen.

Once all eight sets had been divided into groups using this procedure, participants were re-presented with their groupings a set at a time, and were asked to give any group containing two or more members a simple descriptive name. (Participants were also asked to write a few sentences explaining what had led them to classify each named group of descriptions together, though this data will not be analysed here).

The experiment took around an hour to complete.

## Results

For every object description set, the groups formed by each participant's classifications were analysed (with the results displayed in Table 1). Groupings which emerged fell into a number of broad patterns . These classification types are listed in Table 2, below. Similarities across groupings (i.e. similarity shared by every member of a two or more member group across a categorised object description set) which could be identified according to Gentner et al's taxonomy were found in 80% of groupings (in Types 1, 3, 4, 5 and 6).

The most common grouping pattern used was of Type 1 (groups divided into: 1. B-LS-SSO: 2. SMAO-MA: 3. OO, using a network of systematic causal relations), which accounted for 70% of all classified object description sets.

The next largest grouping, comprising 5% of the total was of Type 6. These sets were grouped using a largely structural criterion which resulted in the same grouping pattern as for Type 1 with the exception of the SSO object description which was grouped on its own, even though MA and SMAO were still grouped together.

Groupings which occurred due to participants using common first order relations (those of Type 3) occurred in 1.9% of cases.

Object description sets were grouped according to Types 4 and 5 in 3.1% of cases. The only similarity across groupings of these types is that the object descriptions in each group had only objects in common.

Other groupings worth mentioning were Types 7 and 8, in which the structured object descriptions were grouped according to a determinable pattern, (structure for Type 7, 4.4%) and object attributes (Type 8, 1.9%). but the OO descriptions were assigned according to features in Type 7 (where we would expect a separate grouping), and grouped separately in Type 8 (grouped with descriptions containing similar object attributes expected).

Only 0.6% (one occurrence) of groupings were of Type 2, where the base was put into a category of its own, with shared structure being the only similarity across groupings.

In 11.2% of groupings it was impossible to determine an overall criterion for determining the pattern produced; each of these groupings had only a single occurrence.

**Object Description Sets**

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Type 1 Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | 1 | 1 | 1 | 5 | 1 | 7 | 1 | 7 | 5 |
| | B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| | C | 1 | - | 1 | 1 | 1 | 3 | 1 | 1 | 6 |
| | D | - | 1 | 1 | 1 | 1 | 8 | - | 1 | 5 |
| | E | 1 | 1 | 1 | 1 | 4 | - | 1 | 1 | 6 |
| S | F | 1 | 1 | 1 | 1 | 1 | 1 | - | 7 | 6 |
| u | G | 1 | - | 1 | 1 | 1 | - | 1 | - | 5 |
| b | H | - | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| j | I | 1 | 1 | 1 | 1 | 6 | - | 1 | 1 | 6 |
| e | J | 1 | 1 | 2 | 1 | 1 | - | - | 1 | 5 |
| c | K | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| t | L | 1 | 1 | 1 | 1 | 1 | - | 1 | 1 | 7 |
| s | M | 1 | 6 | 1 | 6 | 1 | 1 | 1 | 8 | 5 |
| | N | 8 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 6 |
| | O | 1 | 1 | 1 | 1 | 1 | 1 | 6 | 6 | 6 |
| | P | 1 | 6 | 1 | 1 | 1 | 7 | 1 | 7 | 5 |
| | Q | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| | R | 1 | 7 | 1 | - | 1 | - | 6 | 7 | 3 |
| | S | 9 | 1 | 1 | 6 | 1 | - | 9 | - | 3 |
| | T | 1 | 4 | 3 | 1 | 3 | - | 4 | 9 | 2 |

**Table 1:** Results for grouping patterns. Each participant was given 8 sets of object descriptions (each row represents one participant; each column an object description set): the type of grouping is indicated by the type number in the object description set column (see also Table 2).

## Discussion

This study further examined the hypothesis put forward by Ramscar and Pain (1996) that categorisation judgements in humans can be determined more by shared structural systematicity than by shared object attributes (surface features) between the objects/ things/ rituals to be classified. The results show considerable evidence to support this hypothesis: 70% of the groupings were made in this way (had participants grouped randomly, mathematical combinatorics yield 213 possible groupings of the materials). In a further 10% of groupings (Types 2, 6 and 7), shared structure was clearly the criterion determining the participants' overall groupings, although a single object description was classified unaccountably (usually singly).

An interesting effect from the Ramscar and Pain study that - intentionally - was not replicated in this experiment, was the production of a large number of Type 2 groupings. In their experiment, Ramscar and Pain left an extra structure (inserted by Gentner et al as part of their analogy study) in a subset of the base stories presented to participants. These base stories with extra structure then tended to be grouped singularly (see Type 2 in Table 2, below). Since the 'Karla the Pot' materials did not contain any extra structures in the Base, we did not expect significant numbers of Type 2 stories to be produced, and in the event, only 0.6% of groupings (1 out of 180) resulted in a Type 2 pattern, where the base was classified singularly in an otherwise structurally determined grouping pattern.

| Classification Criterion | Number | % of Total |
|---|---|---|
| Systematic network of relations in common - Type 1 | | |
| 1 B LS SSO 2 SMAO MA 3 OO | 112 | 70% |
| Systematic network of relations in common - Type 2 (Ramscar & Pain, 1996) (Base classified separately) | | |
| 1 LS SSO 2 SMAO MA 3 B 4 OO | 1 | 0.6% |
| First order relations in common - Type 3 | | |
| 1 B LS SSO SMAO MA 2 OO | 3 | 1.9% |
| Only object similarities in common Types 4 & 5 | | |
| 1 MA LS B OO 2 SSO SMAO | 4 | 2.5% |
| 1 B OO 2 LS MA 3 SSO SMAO | 1 | 0.6% |
| Largely systematic network of relations in common Type 6 | | |
| 1 B LS 2 SSO 3 MA SMAO 4 00 | 8 | 5% |
| OO 'Problems' - Structure based - Type 7; Object attribute based - Type 8 | | |
| 1 B LS SSO OO 2 MA SMAO | 7 | 4.4% |
| 1 B LS MA 2 SSO SMAO 3 OO | 3 | 1.9% |
| Type 9 and others - No clear pattern | | |
| 1 B LS 2 SSO 3 MA 4 SMAO 5 OO | 3 | 1.9% |
| Others | 18 | 11.2% |

Table 2: Output patterns from the categorisation task, showing the groups formed and criteria established. The object descriptions are labelled according to Gentner's taxonomy of similarity (defined above): B = Base; LS = Literal Similarity; SSO = Structural Similarity Only; SMAO = Structural Similarity with MA and Object Similarity with SSO; MA = Mere Appearance; OO = Object Only match.

Groupings that appeared to be formed on the basis of shared surface attributes only amounted to 3.1% of the total (Types 4 & 5). To these could be added another 1.9% of groupings (Type 8) in which shared features determined the overall groupings, although the OO object description - distinctive due to its complete lack of any systematic structure - was classified separately.

Of those object descriptions classified according to shared object attributes, only 2.5% (Type 4 groupings) reflect the 'normal' categories shown in Figure 2.

Clearly, structure appears to be the key determinant of participants' classifications in this study. Typically, categorisation models have tended to concentrate on object descriptions, making use of very representationally-simple attribute-value lists (see Murphy and Medin, 1985; Komatsu, 1992), whereas, analogy research has examined relationships between highly structured representations (considering the influence of attributes, relations and higher-order relations in judgements of similarity. The evidence of this study would appear to support the claim that more notice needs to be taken of the kinds of representations used - and the effects representations produce - in categorisation studies (Medin, Goldstone, and Gentner, 1993; Ramscar and Pain 1996).

The results of the present study also support the broader findings of Ramscar and Pain (1996), who conjectured that the processes underlying analogy and categorisation are not as distinct as is usually proposed. Both their results and ours show shared structural systematicity (Gentner, 1983) as the main process underlying categorisation judgements in the particular experimental conditions. Ordinarily, structural systematicity has been considered the domain of analogy, rather than categorisation.

In this study, the influence of shared structural systematicity has been remarkable. Participants have preferred groupings between pots and walls, and walls and pans, to pots and pans and walls alone. Whilst we feel that these findings have strong implications for categorisation research, we also feel that they should cause some food for thought as regards the way that analogy is typically viewed. As noted earlier, there is a widespread acceptance in analogy research of the two-process view of analogical / metaphorical and literal understandings, whereby 'literal' (within category) understandings are external to non-literal (analogical or metaphorical) understandings, and are therefore assumed to be computed by separate cognitive processes.

The evidence of this study can join other theoretical and empirical evidence against a two-process account of literal (categorical) versus non-literal (analogical or metaphorical) reasoning: we mentioned earlier Hoffman and Kemper's (1987) review of reaction time studies, which convincingly demonstrates the meagre evidence for the widely held belief that literal (intra-categorical) meanings are processed faster than metaphorical or analogical (inter-categorical) meanings.

In spite of this, we do not want to say that analogy *is* categorisation. It is difficult to envisage how such a central cognitive process such as categorisation could be reduced to a single process (c.f. Goldstone, 1994). Given the difficulty inherent in characterising analogical, metaphorical and categorical reasoning (Wittgenstein, 1953; Ramscar, 1997), we are as dubious of the usefulness of the kind of identity statements made by Glucksberg and Keysar, (1990), who argue that metaphorical statements should be understood as class-inclusion statements, as we are of the contrast definitions with which we started this account; we consider it plausible - even likely - that a number of reasoning processes play a part in categorisation. Rather, like Ramscar and Pain (1996) we argue that - in the light of the evidence currently available - analogy is best viewed as a *sub-process* of categorisation, and not as a separate process. Ultimately, we believe that the adherence to the two-process account confuses the relationship between analogy and categorisation and is hampering progress towards further understanding of both (Ramscar, Pain and Cooper, 1997). Until there is a better empirical and theoretical basis to do so, we argue that it may be useful (and more honest) to keep an open mind as to whether an orange crate is an orange crate *can* be a table?

## Acknowledgements

The authors would like to thank Helen Pain and Richard Cooper for their assistance in this work, and Dedre Gentner for providing the "Karla the Hawk" story sets. We are also grateful for the comments of two anonymous reviewers.

## References

Falkenhainer, B., Forbus, K.D. and Gentner, D. (1989) The structure mapping engine: an algorithm and examples. *Artificial Intelligence*, **41**, 1-63

Gentner, D. (1983) Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, **7**: 155-170.

Forbus K, Gentner, D, and Law , K (1995) MAC/FAC: A model of similarity based retrieval *Cognitive Science* **19**:2:141-205

French, RM (1995) *The Subtlety of Sameness*, MIT Press, Cambridge, Ma.

Gentner, D Ratterman, M.J. and Forbus, K. (1993) The roles of similarity in transfer. *Cognitive Psychology* **25**: 524-575

Goldstone, R.L. (1994) The role of similarity in categorization: providing a groundwork. *Cognition*, **52**: 125-157

Glucksberg, S. and Keysar, B. (1990) Understanding metaphorical comparisons: Beyond Similarity. *Psychological Review*, **97**: 1: 3-18

Hoffman, R R and Kemper, S (1987) What could reaction-time studies be telling us about metaphor comprehension? *Metaphor and Symbolic Activity* **2(3)**, 149 - 186

Holyoak, K.J. and Thagard, P. (1995) *Mental Leaps*. MIT Press, Cambridge, Ma.

Komatsu, LK (1992) Recent views of conceptual structure. *Psychological Bulletin*, **112(3)**, 500-526

Medin, D., Goldstone, R. and Gentner, D. (1993) Respects for similarity. *Psychological Review*, **100**:2:254-278.

Murphy, G.L. and Medin, D.L. (1985) The role of theories in conceptual coherence. *Psychological review*, **92**, 289 - 316

Ramscar, MJA & Pain, HG (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? (*In Proc.*) *18th Annual Conference of the Cognitive Science Society*. Lawrence Earlbaum Assoc., New Jersey.

Ramscar, MJA, Pain, HG & Cooper, R (1997) Is there place for semantic similarity in the analogical mappi process? (*In Proc.*) *19th Annual Conference of Cognitive Science Society*. Lawrence Earlbaum Assc New Jersey.

Ramscar, MJA (1997) Wittgenstein and the nature psychological categories *Similarity and categorisati proceedings of SimCat 97*, Department of Artific Intelligence Conference Proceedings, University Edinburgh, Scotland, 205-211.

Récanati, F (1995) The alleged priority of lite interpretation *Cognitive Science* **19**:207-273

Rosch, E. (1978) *Cognition and Categorisation*. Lawre Earlbaum Assoc., New Jersey.

Wittgenstein, L *translated by* Anscombe, E (195 *Philosophical Investigations* Blackwell, Oxford.